Chapter 2
Constant Envelope Modulations

2.1 The Need for Constant Envelope

Digital communication systems operate in the presence of path loss and atmospheric-induced fading. In order to maintain sufficient received power at the destination, it is required that a device for generating adequate transmitter output power based on fixed- but-limited available power be employed, examples of which are traveling-wave tube amplifiers (TWTAs) and solid-state power amplifiers (SSPAs) operated in full- saturation mode to maximize conversion efficiency. Unfortunately, this requirement introduces amplitude modulation-amplitude modulation (AM-AM) and amplitude modulation-phase modulation (AM-PM) conversions into the transmitted signal. Because of this, modulations that transmit information via their amplitude, e.g., quadrature amplitude modulation (QAM), and therefore need a linear amplifying characteristic, are not suitable for use on channels operated in the above maximum transmitter power efficiency requirement.[1] Another consideration regarding radio frequency (RF) amplifier devices that operate in a nonlinear mode at or near saturation is the spectral spreading that they reintroduce due to the nonlinearity subsequent to bandlimiting the modulation prior to amplification. Because of the need for the transmitted power spectrum to fall under a specified mask imposed by regulating agencies such as the FCC or International Telecommunications Union (ITU), the modulation must be designed to keep this spectral spreading to a minimum. This constraint necessitates limiting the amount of instantaneous amplitude fluctuation in the transmitted waveform in addition to imposing the requirement for constant envelope.

[1] An approach whereby it might be possible to generate QAM-type modulations using separate nonlinearly operated high-power amplifiers on the inphase (I) and quadrature (Q) channels is currently under investigation by the author.

Because of the above considerations regarding the need for high transmitter power efficiency, it is clearly desirable to consider modulations that achieve their bandwidth efficiency by means other than resorting to multilevel amplitude modulation. Such constant envelope modulations are the subject of discussion in the first part of this monograph. Because of the large number of possible candidates, to keep within the confines of a reasonable size book, we shall restrict our attention to only those that have some form of inphase-quadrature phase (I-Q) representation and as such an I-Q form of receiver.

2.2 Quadriphase-Shift-Keying and Offset (Staggered) Quadriphase-Shift-Keying

M-ary phase-shift-keying (M-PSK) produces a constant envelope signal that is mathematically modeled in complex form[2] as

$$\tilde{s}\,(t) = \sqrt{2P}e^{j(2\pi f_c t+\theta(t)+\theta_c)} = \tilde{S}\,(t)\,e^{j(2\pi f_c t+\theta_c)} \tag{2.2-1}$$

where P is the transmitted power, f_c is the carrier frequency in hertz, θ_c is the carrier phase, and $\theta(t)$ is the data phase that takes on equiprobable values $\beta_i = (2i-1)\,\pi/M,\ i = 1, 2, \cdots, M$, in each symbol interval, T_s. As such, $\theta(t)$ is modeled as a random pulse stream, that is,

$$\theta(t) = \sum_{n=-\infty}^{\infty} \theta_n p\,(t - nT_s) \tag{2.2-2}$$

where θ_n is the information phase in the nth symbol interval, $nT_s < t \leq (n+1)T_s$, ranging over the set of M possible values β_i as above, and $p(t)$ is a unit amplitude rectangular pulse of duration T_s seconds. The symbol time, T_s, is related to the bit time, T_b, by $T_s = T_b \log_2 M$ and, thus, the nominal gain in bandwidth efficiency relative to binary phase-shift-keying (BPSK), i.e., $M = 2$, is a factor of $\log_2 M$. The signal constellation is a unit circle with points uniformly spaced by $2\pi/M$ rad. Thus, the complex signal transmitted in the nth symbol interval is

$$\tilde{s}(t) = \sqrt{2P}e^{j(2\pi f_c t+\theta_n+\theta_c)}, \quad nT_s < t \leq (n+1)T_s, \quad n = -\infty, \cdots, \infty \tag{2.2-3}$$

[2] The actual (real) transmitted signal is $s\,(t) = \mathrm{Re}\,\{\tilde{s}(t)\} = \sqrt{2P}\cos\left(2\pi f_c t + \theta(t) + \theta_c\right)$.

Note that because of the assumed rectangular pulse shape, the complex baseband signal $\tilde{S}(t) = \sqrt{2P}e^{j\theta_n}$ is constant in this same interval and has envelope $\left|\tilde{S}(t)\right| = \sqrt{2P}$.

A special case of M-PSK that has an I-Q representation is quadriphase-shift-keying (QPSK), and corresponds to $M = 4$. Here it is conventional to assume that the phase set $\{\beta_i\}$ takes on values $\pi/4,\ 3\pi/4, 5\pi/4, 7\pi/4$. Projecting these information phases on the quadrature amplitude axes, we can equivalently write QPSK in the nth symbol interval in the complex I-Q form[3]

$$\tilde{s}(t) = \sqrt{P}\left(a_{In} + ja_{Qn}\right)e^{j(2\pi f_c t + \theta_c)}, \quad nT_s < t \leq (n+1)T_s \tag{2.2-4}$$

where the information amplitudes a_{In} and a_{Qn} range independently over the equiprobable values ± 1. Here again, because of the assumed rectangular pulse shape, the complex baseband signal $\tilde{S}(t) = \sqrt{P}\left(a_{In} + ja_{Qn}\right)$ is constant in this same interval. The real transmitted signal corresponding to (2.2-4) has the form

$$s(t) = \sqrt{P}m_I(t)\cos(2\pi f_c t + \theta_c) - \sqrt{P}m_Q(t)\sin(2\pi f_c t + \theta_c),$$

$$m_I(t) = \sum_{n=-\infty}^{\infty} a_{In}p(t - nT_s), \quad m_Q(t) = \sum_{n=-\infty}^{\infty} a_{Qn}p(t - nT_s) \tag{2.2-5}$$

If one examines the form of (2.2-4) it becomes apparent that a large fluctuation of the instantaneous amplitude between symbols corresponding to a 180-deg phase reversal can occur when both a_{In} and a_{Qn} change polarity at the same time. As mentioned in Sec. 2.1, it is desirable to limit the degree of such fluctuation to reduce spectral regrowth brought about by the transmit amplifier nonlinearity, i.e., the smaller the fluctuation, the smaller the sidelobe regeneration and vice versa. By offsetting (staggering) the I and Q modulations by $T_s/2$ s, one guarantees the fact that a_{In} and a_{Qn} cannot change polarity at the same time. Thus, the maximum fluctuation in instantaneous amplitude is now limited to that corresponding to a 90-deg phase reversal (i.e., either a_{In} or a_{Qn}, but not both, change polarity). The resulting modulation, called offset (staggered) QPSK (OQPSK), has a signal of the form

[3] One can think of the complex carrier as being modulated now by a complex random pulse stream, namely, $\tilde{a}(t) = \sum_{n=-\infty}^{\infty}\left(a_{In} + ja_{Qn}\right)p(t - nT_s)$.

$$s(t) = \sqrt{P} m_I(t) \cos(2\pi f_c t + \theta_c) - \sqrt{P} m_Q(t) \sin(2\pi f_c t + \theta_c),$$

$$m_I(t) = \sum_{n=-\infty}^{\infty} a_{In} p(t - nT_s), \quad m_Q(t) = \sum_{n=-\infty}^{\infty} a_{Qn} p\left(t - \left(n + \frac{1}{2}\right) T_s\right) \tag{2.2-6}$$

While it is true that for M-PSK with $M = 2^m$ and m an arbitrary integer, the information phases can be projected on the I and Q coordinates and as such obtain, in principle, an I-Q transmitter representation, it should be noted that the number of possible I-Q amplitude pairs obtained from these projections exceeds M. Consequently, decisions on the resulting I and Q multilevel amplitude signals at the receiver are not independent in that each pair of amplitude decisions does not necessarily render one of the transmitted phases. Therefore, for $M \geq 8$ it is not practical to view M-PSK in an I-Q form.

The detection of an information phase can be obtained by combining the detections on the I and Q components of this phase. The receiver for QPSK is illustrated in Fig. 2-1(a) while the analogous receiver for OQPSK is illustrated in Fig. 2-1(b). The decision variables that are input to the hard-limiting threshold devices are

$$\left.\begin{aligned} y_{In} &= a_{In}\sqrt{P}T_s + N_{In} \\ y_{Qn} &= a_{Qn}\sqrt{P}T_s + N_{Qn} \end{aligned}\right\} \tag{2.2-7}$$

where for QPSK

$$\left.\begin{aligned} N_{In} &= \mathrm{Re}\left\{\int_{nT_s}^{(n+1)T_s} \tilde{N}(t)\, dt\right\} \\ N_{Qn} &= \mathrm{Im}\left\{\int_{nT_s}^{(n+1)T_s} \tilde{N}(t)\, dt\right\} \end{aligned}\right\} \tag{2.2-8}$$

whereas for OQPSK

$$\left.\begin{aligned} N_{In} &= \mathrm{Re}\left\{\int_{nT_s}^{(n+1)T_s} \tilde{N}(t)\,dt\right\} \\ N_{Qn} &= \mathrm{Im}\left\{\int_{(n+1/2)T_s}^{(n+3/2)T_s} \tilde{N}(t)\,dt\right\} \end{aligned}\right\} \tag{2.2-9}$$

In either case, N_{In}, N_{Qn} are zero mean Gaussian random variables (RVs) with variance $\sigma_N^2 = N_0 T_s/2$ and thus conditioned on the data symbols, y_{In}, y_{Qn} are also Gaussian RVs with the same variance.

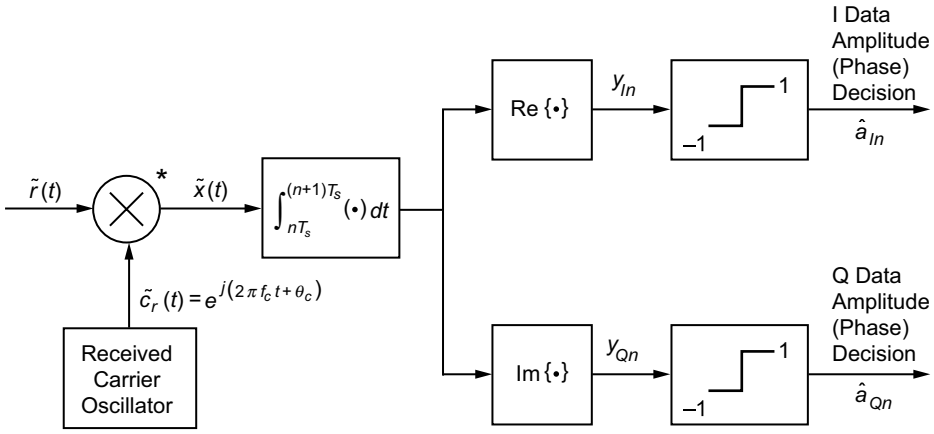

Fig. 2-1(a). Complex form of optimum receiver for ideal coherent detection of QPSK over the AWGN.

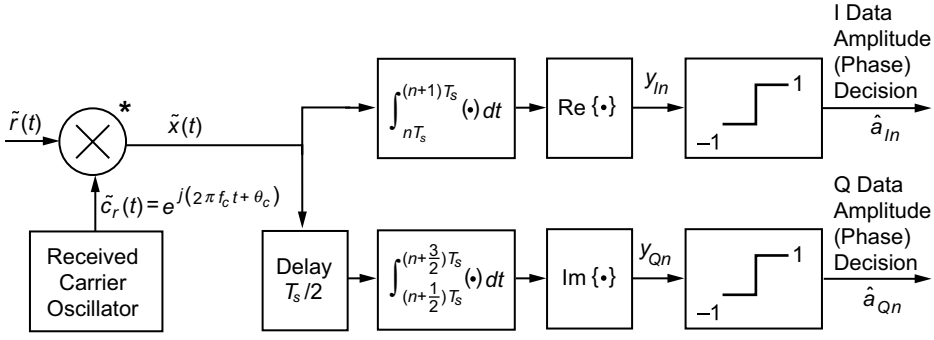

Fig. 2-1(b). Complex form of optimum receiver for ideal coherent detection of OQPSK over the AWGN.

2.3 Differentially Encoded QPSK and Offset (Staggered) QPSK

In an actual coherent communication system transmitting M-PSK modulation, means must be provided at the receiver for establishing the local demodulation carrier reference signal. This means is traditionally accomplished with the aid of a suppressed carrier-tracking loop [1, Chap. 2]. Such a loop for M-PSK modulation exhibits an M-fold phase ambiguity in that it can lock with equal probability at the transmitted carrier phase plus any of the M information phase values. Hence, the carrier phase used for demodulation can take on any of these same M phase values, namely, $\theta_c + \beta_i = \theta_c + 2i\pi/M,\ i = 0, 1, 2, \cdots, M-1$. Coherent detection cannot be successful unless this M-fold phase ambiguity is resolved.

One means for resolving this ambiguity is to employ differential phase encoding (most often simply called differential encoding) at the transmitter and differential phase decoding (most often simply called differential decoding) at the receiver following coherent detection. That is, the information phase to be communicated is modulated on the carrier as the difference between two adjacent transmitted phases, and the receiver takes the difference of two adjacent phase decisions to arrive at the decision on the information phase.[4] In mathematical terms, if $\Delta\theta_n$ were the information phase to be communicated in the nth transmission interval, the transmitter would first form $\theta_n = \theta_{n-1} + \Delta\theta_n$ modulo 2π (the differential encoder) and then modulate θ_n on the carrier.[5] At the receiver, successive decisions on θ_{n-1} and θ_n would be made and then differenced modulo 2π (the differential decoder) to give the decision on $\Delta\theta_n$. Since the decision on the true information phase is obtained from the difference of two adjacent phase decisions, a performance penalty is associated with the inclusion of differential encoding/decoding in the system.

For QPSK or OQPSK, the differential encoding/decoding process can be performed on each of the I and Q channels independently. A block diagram of a receiver for differentially encoded QPSK (or OQPSK) would be identical to that shown in Fig. 2-1(a) [or Fig. 2-1(b)], with the inclusion of a binary differential decoder in each of the I and Q arms following the hard-decision devices [see

---

[4] Note that this receiver (i.e., the one that makes optimum coherent decisions on two successive symbol phases and then differences these to arrive at the decision on the information phase) is suboptimum when $M > 2$ [2]. However, this receiver structure, which is the one classically used for coherent detection of differentially encoded M-PSK, can be arrived at by a suitable approximation of the likelihood function used to derive the true optimum receiver, and at high signal-to-noise ratio (SNR), the difference between the two becomes mute.

[5] Note that we have shifted our notation here insofar as the information phases are concerned so as to keep the same notation for the phases actually transmitted.

Figs. 2-2(a) and 2-2(b)].[6] Inclusion of differentially encoded OQPSK in our discussion is important since, as we shall see later on, other forms of modulation, e.g., minimum-shift-keying (MSK), have an I-Q representation in the form of pulse-shaped, differentially encoded OQPSK.

2.4 π/4-QPSK: A Variation of Differentially Encoded QPSK with Instantaneous Amplitude Fluctuation Halfway between That of QPSK and OQPSK

Depending on the set of phases, $\{\Delta\beta_i\}$, used to represent the information phase, $\Delta\theta_n$, in the nth transmission interval, the actual transmitted phase, θ_n, in this same transmission interval can range either over the same set, $\{\beta_i\} = \{\Delta\beta_i\}$, or over another phase set. If for QPSK, we choose the set $\Delta\beta_i = 0, \pi/2, \ \pi, 3\pi/2$ to represent the information phases, then starting with an initial transmitted phase chosen from the set $\pi/4, \ 3\pi/4, 5\pi/4, 7\pi/4$, the subsequent transmitted phases, $\{\theta_n\}$, will also range over the set $\pi/4, \ 3\pi/4, 5\pi/4, 7\pi/4$ in every transmission interval. This is the conventional form of differentially encoded QPSK, as discussed in the previous section. Now suppose instead that the set $\Delta\beta_i = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$ is used to represent the information phases, $\{\Delta\theta_n\}$. Then, starting, for example, with an initial phase chosen from the set $\pi/4, 3\pi/4, 5\pi/4, 7\pi/4$, the transmitted phase in the next interval will range over the set $0, \pi/2, \pi, 3\pi/2$. In the following interval, the transmitted phase will range over the set $\pi/4, 3\pi/4, 5\pi/4, 7\pi/4$, and in the interval following that one, the transmitted phase will once again range over the set $0, \pi/2, \pi, 3\pi/2$. Thus, we see that for this choice of phase set corresponding to the information phases, $\{\Delta\theta_n\}$, the transmitted phases, $\{\theta_n\}$, will alternatively range over the sets $0, \pi/2, \pi, 3\pi/2$ and $\pi/4, 3\pi/4, 5\pi/4, 7\pi/4$. Such a modulation scheme, referred to as $\pi/4$-QPSK [3], has an advantage relative to conventional differentially encoded QPSK in that the maximum change in phase from transmission to transmission is 135 deg, which is halfway between the 90-deg maximum phase change of OQPSK and 180-deg maximum phase change of QPSK.

In summary, on a linear additive white Gaussian noise (AWGN) channel with ideal coherent detection, all three types of differentially encoded QPSK, i.e., conventional (nonoffset), offset, and $\pi/4$ perform identically. The differences among the three types on a linear AWGN channel occur when the carrier demodulation phase reference is not perfect, which corresponds to nonideal coherent detection.

---

[6] Since the introduction of a 180-deg phase shift to a binary phase sequence is equivalent to a reversal of the polarity of the binary data bits, a binary differential encoder is characterized by $a_n = a_{n-1}b_n$ and the corresponding binary differential decoder is characterized by $b_n = a_{n-1}a_n$ where $\{b_n\}$ are now the information bits and $\{a_n\}$ are the actual transmitted bits on each channel.

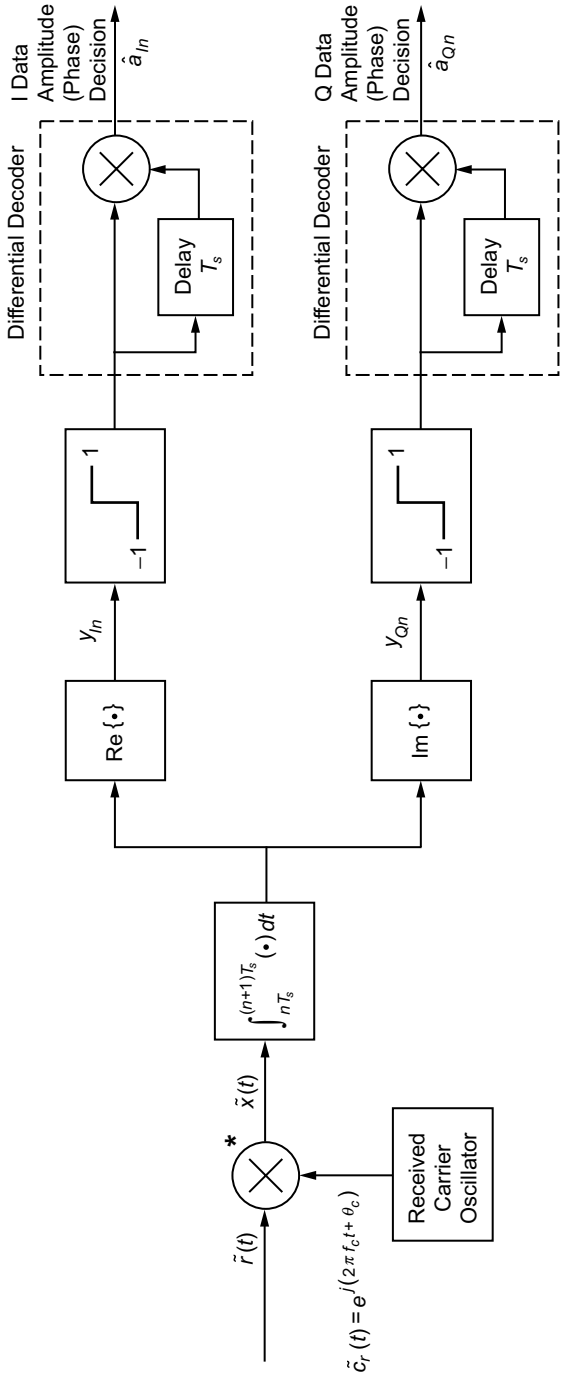

Fig. 2-2(a). Complex form of optimum receiver for ideal coherent detection of differentially encoded QPSK over the AWGN.

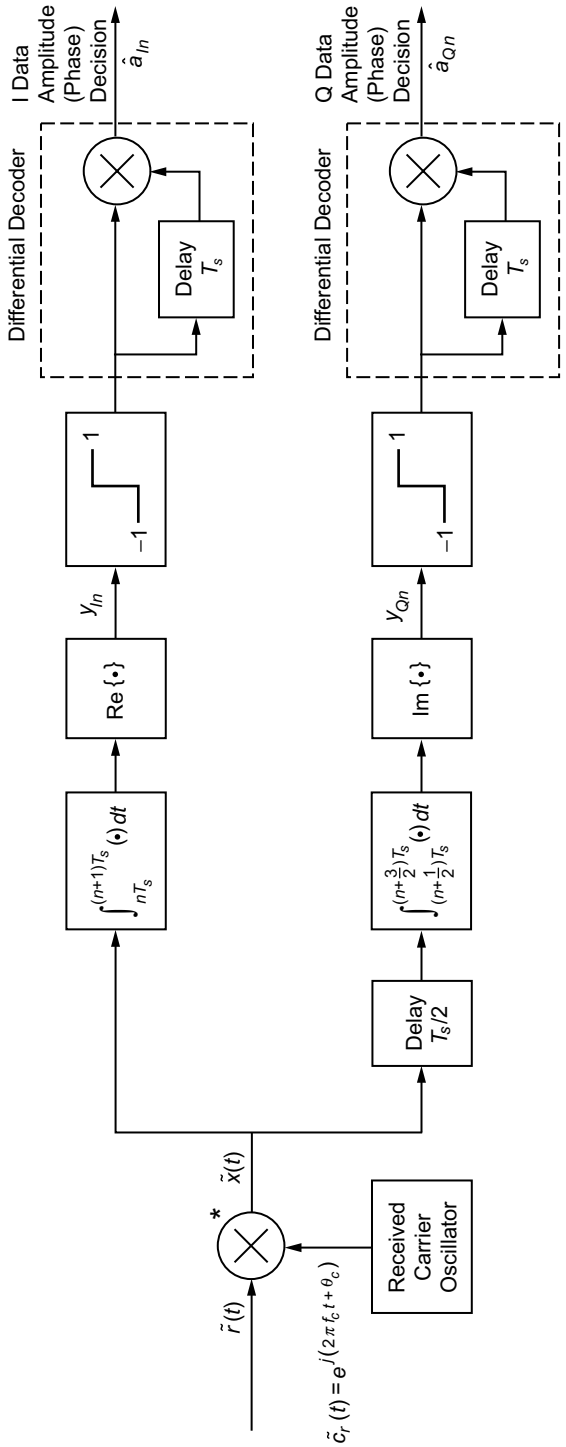

Fig. 2-2(b). Complex form of optimum receiver for ideal coherent detection of differentially encoded OQPSK over the AWGN.

2.5 Power Spectral Density Considerations

The power spectral densities (PSD) of QPSK, OQPSK, and the differentially encoded versions of these are all identical and are given by

$$S(f) = PT_s \left(\frac{\sin \pi f T_s}{\pi f T_s} \right)^2 \tag{2.5-1}$$

We see that the asymptotic (large f) rate of rolloff of the PSD varies as f^{-2}, and a first null (width of the main lobe) occurs at $f = 1/T_s = 1/2T_b$. Furthermore, when compared with BPSK, QPSK is exactly twice as bandwidth efficient.

2.6 Ideal Receiver Performance

Based upon the decision variables in (2.2-7) the receiver for QPSK or OQPSK makes its I and Q data decisions from

$$\left.\begin{aligned} \hat{a}_{In} &= \operatorname{sgn} y_{In} \\ \hat{a}_{Qn} &= \operatorname{sgn} y_{Qn} \end{aligned}\right\} \tag{2.6-1}$$

which results in an average bit-error probability (BEP) given by

$$P_b(E) = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{E_b}{N_0}} \right), \quad E_b = PT_b \tag{2.6-2}$$

and is identical to that of BPSK. Thus, we conclude that ideally BPSK, QPSK, and OQPSK have the identical BEP performance although the latter two occupy half the bandwidth.

2.7 Performance in the Presence of Nonideal Transmitters

2.7.1 Modulator Imbalance and Amplifier Nonlinearity

The deleterious effect on receiver performance of modulator phase and amplitude imbalance and amplifier nonlinearity has been studied by several researchers [3–10]. With regard to modulator imbalances, the primary source of degradation comes about because of the effect of the imbalance on the steady-state lock point of the carrier tracking loop, which has a direct impact on the determination of

accurate average BEP performance. Here, we summarize some of these results for QPSK and OQPSK, starting with modulator imbalance acting alone and then later on in combination with amplifier nonlinearity. We begin our discussion with a description of an imbalance model associated with a modulator for generating these signals.

2.7.1.1 Modulator Imbalance Model. QPSK can be implemented with two balanced modulators, one on each of the I and Q channels, as illustrated in Fig. 2-3. Each of these modulators is composed of two AM modulators with inputs equal to the input nonreturn-to-zero (NRZ) data stream and its inverse (bit polarities inverted). The difference of the outputs of the two AM modulators serves as the BPSK transmitted signal on each channel. A mathematical description of the I and Q channel signals in the presence of amplitude and phase imbalances introduced by the AM modulators is[7]

$$s_I(t) = \frac{\sqrt{P}}{2} m_I(t) \left[\cos(2\pi f_c t + \theta_{cI}) + \Gamma_I \cos(2\pi f_c t + \theta_{cI} + \Delta\theta_{cI})\right] + \frac{\sqrt{P}}{2}\left[\cos(2\pi f_c t + \theta_{cI}) - \Gamma_I \cos(2\pi f_c t + \theta_{cI} + \Delta\theta_{cI})\right] \quad (2.7\text{-}1a)$$

$$s_Q(t) = \frac{\sqrt{P}}{2} m_Q(t) \left[\sin(2\pi f_c t + \theta_{cQ}) + \Gamma_Q \sin(2\pi f_c t + \theta_{cQ} + \Delta\theta_{cQ})\right] + \frac{\sqrt{P}}{2}\left[\sin(2\pi f_c t + \theta_{cQ}) - \Gamma_Q \sin(2\pi f_c t + \theta_{cQ} + \Delta\theta_{cQ})\right] \quad (2.7\text{-}1b)$$

$$s(t) = s_I(t) + s_Q(t)$$

where θ_{cI}, θ_{cQ} are the local oscillator carrier phases associated with the I and Q balanced modulators, Γ_I, Γ_Q (both assumed to be less than unity) are the relative amplitude imbalances of these same modulators, and $\Delta\theta_{cI}, \Delta\theta_{cQ}$ are the phase imbalances between the two AM modulators in each of the I and Q

[7] To be consistent with the usage in Ref. 8, we define the transmitted signal as the sum of the I and Q signals, i.e., $s(t) = s_I(t) + s_Q(t)$ rather than their difference as in the more traditional usage of (2.2-5). This minor switch in notation is of no consequence to the results that follow.

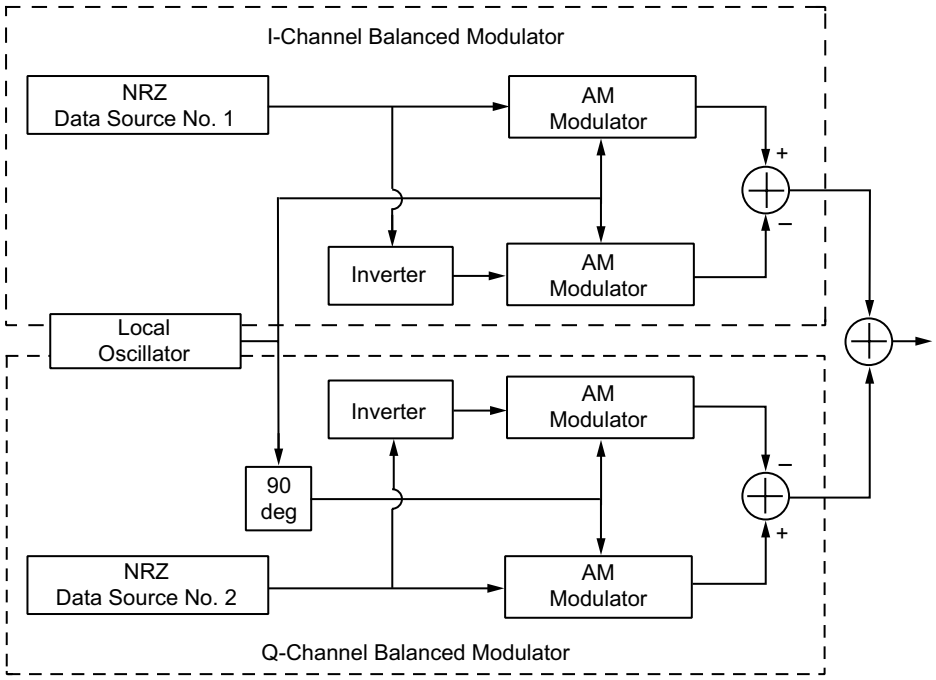

Fig. 2-3. Balanced QPSK modulator implementation.

balanced modulators, respectively. Note that by virtue of the fact that we have introduced separate notation for the I and Q local oscillator phases, i.e., θ_{cI} and θ_{cQ}, we are also allowing for other than a perfect 90-deg phase shift between I and Q channels. Alternatively, the model includes the possibility of an interchannel phase imbalance, $\Delta\theta_c = \theta_{cI} - \theta_{cQ}$. Since we will be interested only in the difference $\Delta\theta_c$, without loss of generality we shall assume $\theta_{cQ} = 0$, in which case $\theta_{cI} = \Delta\theta_c$. Finally, note that if $\Gamma_I = \Gamma_Q = 1$, $\Delta\theta_{cI} = \Delta\theta_{cQ} = 0$, and $\theta_{cI} = \theta_{cQ} = \theta_c$, then we obtain balanced QPSK as characterized by (2.2-5).

As shown in Ref. 8, the transmitted signal of (2.7-1a) and (2.7-1b) can, after some trigonometric manipulation, be written in the form

$$s(t) = \sqrt{P}\Big\{\big[\alpha_I + \beta_I m_I(t) - \gamma_Q\big(1 - m_Q(t)\big)\big]\cos 2\pi f_c t$$

$$+\big[\alpha_Q + \beta_Q m_Q(t) + \delta_I - \gamma_I m_I(t)\big]\sin 2\pi f_c t\Big\} \tag{2.7-2}$$

where

$$\left.\begin{aligned}
\alpha_I &= \frac{(1-\Gamma_I\cos\Delta\theta_{cI})\cos\Delta\theta_c + \Gamma_I\sin\Delta\theta_{cI}\sin\Delta\theta_c}{2},\\
\alpha_Q &= \frac{1-\Gamma_Q\cos\Delta\theta_{cQ}}{2}\\
\beta_I &= \frac{(1+\Gamma_I\cos\Delta\theta_{cI})\cos\Delta\theta_c - \Gamma_I\sin\Delta\theta_{cI}\sin\Delta\theta_c}{2},\\
\beta_Q &= \frac{1+\Gamma_Q\cos\Delta\theta_{cQ}}{2}\\
\gamma_I &= \frac{(1+\Gamma_I\cos\Delta\theta_{cI})\sin\Delta\theta_c + \Gamma_I\sin\Delta\theta_{cI}\cos\Delta\theta_c}{2},\\
\gamma_Q &= \frac{\Gamma_Q\sin\Delta\theta_{cQ}}{2}\\
\delta_I &= \frac{-(1-\Gamma_I\cos\Delta\theta_{cI})\sin\Delta\theta_c + \Gamma_I\sin\Delta\theta_{cI}\cos\Delta\theta_c}{2}
\end{aligned}\right\} \qquad (2.7\text{-}3)$$

The form of the transmitted signal in (2.7-2) clearly identifies the crosstalk introduced by the modulator imbalances, i.e., the dependence of the I channel signal on the Q channel modulation and vice versa, as well as the lack of perfect quadrature between I and Q channels. Note the presence of a spurious carrier component in (2.7-3), i.e., a discrete (unmodulated) carrier component that is not present in the balanced case. Note that for perfect quadrature between the I and Q channels, i.e., $\Delta\theta_c = 0$, we have $\gamma_I = \delta_I = (1/2)\Gamma_I \sin\Delta\theta_{cI}$, and (2.7-2) becomes the symmetric form

$$\begin{aligned}
s(t) = \sqrt{P}\Big\{ &\left[\alpha_I + \beta_I m_I(t) - \gamma_Q\big(1 - m_Q(t)\big)\right]\cos 2\pi f_c t\\
&+ \left[\alpha_Q + \beta_Q m_Q(t) + \gamma_I\big(1 - m_I(t)\big)\right]\sin 2\pi f_c t\Big\}
\end{aligned} \qquad (2.7\text{-}4)$$

which corresponds to the case of modulator imbalance alone. If now the phase imbalance is removed, i.e., $\Delta\theta_{cI} = \Delta\theta_{cQ} = 0$, then $\gamma_I = \gamma_Q = 0$, and the crosstalk in the transmitted signal disappears, i.e., modulator amplitude imbalance alone does not cause crosstalk. It is important to note, however, that the lack of crosstalk in the transmitted signal does not guarantee the absence

of crosstalk at the receiver, which affects the system error probability performance. Finally, note that for the perfectly balanced case, $\beta_I = \beta_Q = 1$ and $\alpha_I = \alpha_I = 0$, $\gamma_I = \gamma_Q = 0$, and (2.7-4) results in (2.2-5) with the exception of the minus sign discussed in Footnote 7.

2.7.1.2 Effect on Carrier Tracking Loop Steady-State Lock Point. When a Costas-type loop is used to track a QPSK signal, it forms its error signal from $IQ\left(I^2 - Q^2\right)$, where the letters I and Q now refer to signals that are synonymous with the outputs of the inphase and quadrature integrate-and-dump (I&D) filters, y_{In} and y_{Qn}, shown in Fig. 2-2(a). In the presence of modulator imbalance and imperfect I and Q quadrature, the evaluation of the steady-state lock point of the loop was considered in Ref. 8 and, in the most general case, was determined numerically. For the special case of identically imbalanced I and Q modulators and no quadrature imperfection, i.e., $\Gamma_I = \Gamma_Q = \Gamma$, $\Delta\theta_{cI} = \Delta\theta_{cQ} = \Delta\theta_u$ and $\Delta\theta_c = 0$, a closed-form result for the steady-state lock point is possible and is given by

$$\phi_0 = -\frac{1}{4}\tan^{-1}\frac{6\Gamma^2\sin 2\Delta\theta_u + \Gamma^4\sin 4\Delta\theta_u}{1 + 6\Gamma^2\cos 2\Delta\theta_u + \Gamma^4\cos 4\Delta\theta_u} \tag{2.7-5}$$

Note that for perfect modulator amplitude balance ($\Gamma = 1$), we obtain $\phi_0 = -\Delta\theta_u/2$, as expected. This shift in the lock point exists independently of the loop SNR and thus can be referred to as an irreducible carrier phase error.

2.7.1.3 Effect on Average BEP. Assuming that the phase error is constant over the bit time (equivalently, the loop bandwidth is small compared to the data rate) and that the 90-deg phase ambiguity associated with the QPSK Costas loop can be perfectly resolved (e.g., by differential encoding), the average BEP can be evaluated by averaging the conditional (on the phase error, ϕ) BEP over the probability density function (PDF) of the phase error, i.e.,

$$\left.\begin{aligned} P_{bI}(E) &= \int_{\phi_0-\pi/4}^{\phi_0-\pi/4} P_{bI}(E;\phi)\, p_\phi(\phi)\, d\phi \\ P_{bQ}(E) &= \int_{\phi_0-\pi/4}^{\phi_0-\pi/4} P_{bQ}(E;\phi)\, p_\phi(\phi)\, d\phi \end{aligned}\right\} \tag{2.7-6}$$

where

$$p_\phi(\phi) = 4\frac{\exp\Big(\rho_{4\phi}\cos\big(4(\phi-\phi_0)\big)\Big)}{2\pi I_0(\rho_{4\phi})}, \quad |\phi-\phi_0| \le \frac{\pi}{4} \tag{2.7-7}$$

is the usual Tikhonov model assumed for the phase error PDF [11] with ϕ_0 determined from (2.7-5). The parameter $\rho_{4\phi}$ is the loop SNR of the four times phase error process (which is what the loop tracks) and $I_0(\cdot)$ is the modified first-order Bessel function of the first kind. Based on the hard decisions made on y_{In} and y_{Qn} in Fig. 2-2(a), the conditional BEPs on the I and Q channels in the presence of imbalance are given, respectively, in Ref. 8, Eqs. (11a) and (11b):

$$\begin{aligned} P_{bI}(E;\phi) = {} & \frac{1}{8}\mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\cos(\phi+\Delta\theta_c)+\sin\phi\right]\right) \\ & + \frac{1}{8}\mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\cos(\phi+\Delta\theta_c)-\Gamma_Q\sin(\phi+\Delta\theta_{cQ})\right]\right) \\ & + \frac{1}{8}\mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma_I\cos(\phi+\Delta\theta_{c_I}+\Delta\theta_c)-\sin\phi\right]\right) \\ & + \frac{1}{8}\mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma_I\cos(\phi+\Delta\theta_{c_I}+\Delta\theta_c)+\Gamma_Q\sin(\phi+\Delta\theta_{cQ})\right]\right) \end{aligned} \tag{2.7-8a}$$

and

$$\begin{aligned} P_{bQ}(E;\phi) = {} & \frac{1}{8}\mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\cos\phi-\sin(\phi+\Delta\theta_c)\right]\right) \\ & + \frac{1}{8}\mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\cos\phi+\Gamma_I\sin(\phi+\Delta\theta_{cI}+\Delta\theta_c)\right]\right) \\ & + \frac{1}{8}\mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma_Q\cos(\phi+\Delta\theta_{cQ})+\sin(\phi+\Delta\theta_c)\right]\right) \\ & + \frac{1}{8}\mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma_Q\cos(\phi+\Delta\theta_{cQ})-\Gamma_I\sin(\phi+\Delta\theta_{cI}+\Delta\theta_c)\right]\right) \end{aligned} \tag{2.7-8b}$$

Substituting (2.7-7) together with (2.7-8a) and (2.7-8b) in (2.7-6) gives the desired average BEP of the I and Q channels for any degree of modulator imbalance. Note that, in general, the error probability performances of the I and Q channels are not identical.

For a maximum amplitude imbalance (Γ_I or Γ_Q) of 0.2 dB, a maximum phase imbalance ($\Delta\theta_{cI}$ or $\Delta\theta_{cQ}$) of +2 deg, and a maximum I-Q quadrature imbalance ($\Delta\theta_c$) of +2 deg (the values recommended by the CCSDS), Figs. 2-4(a) and 2-4(b) plot the I and Q average BEPs as computed from (2.7-6) for the best and worst combinations of imbalance conditions. In these plots, the loop SNR, $\rho_{4\phi}$, is assumed to have infinite value ("perfect" carrier synchronization), and, consequently, the degradation corresponds only to the shift in the lock point. The case of perfectly balanced QPSK is also included in these plots for comparison purposes. We observe that the best imbalance condition gives a performance virtually identical to that of balanced QPSK, whereas the worst imbalance condition results in an E_b/N_0 loss of 0.33 dB at an average BEP of 10^{-2}.

The extension of the above results to the case of OQPSK is presented in Ref. 9. The same modulator imbalance model as that illustrated in Fig. 2-3 is considered, with the exception that the Q channel data stream is now offset with respect to the I channel data stream, requiring a half-symbol delay between the NRZ data source 2 and AM modulator. Also, the amplitude imbalance, Γ, between the I and Q channels, is now *explicitly* included as an additional independent parameter. Therefore, analogous to (2.7-1b), the Q component of the transmitted OQPSK signal becomes [the I component is still given by (2.7-1a)]

$$s_Q(t) = \Gamma\frac{\sqrt{P}}{2}m_Q\left(t-\frac{T_s}{2}\right)\left[\sin\left(2\pi f_c t+\theta_{cQ}\right)+\Gamma_Q\sin\left(2\pi f_c t+\theta_{cQ}+\Delta\theta_{cQ}\right)\right]$$

$$+\Gamma\frac{\sqrt{P}}{2}\left[\sin\left(2\pi f_c t+\theta_{cQ}\right)-\Gamma_Q\sin\left(2\pi f_c t+\theta_{cQ}+\Delta\theta_{cQ}\right)\right] \qquad (2.7\text{-}9)$$

Using similar trigonometric manipulations for arriving at (2.7-2), the transmitted signal $(s_I(t)+s_Q(t))$ can now be written as

$$s(t)=\sqrt{P}\left\{\left[\alpha_I+\beta_I m_I(t)-\gamma_Q\left(1-m_Q\left(t-\frac{T_s}{2}\right)\right)\right]\cos 2\pi f_c t\right.$$

$$\left.+\left[\alpha_Q+\beta_Q m_Q\left(t-\frac{T_s}{2}\right)+\delta_I-\gamma_I m_I(t)\right]\sin 2\pi f_c t\right\} \qquad (2.7\text{-}10)$$

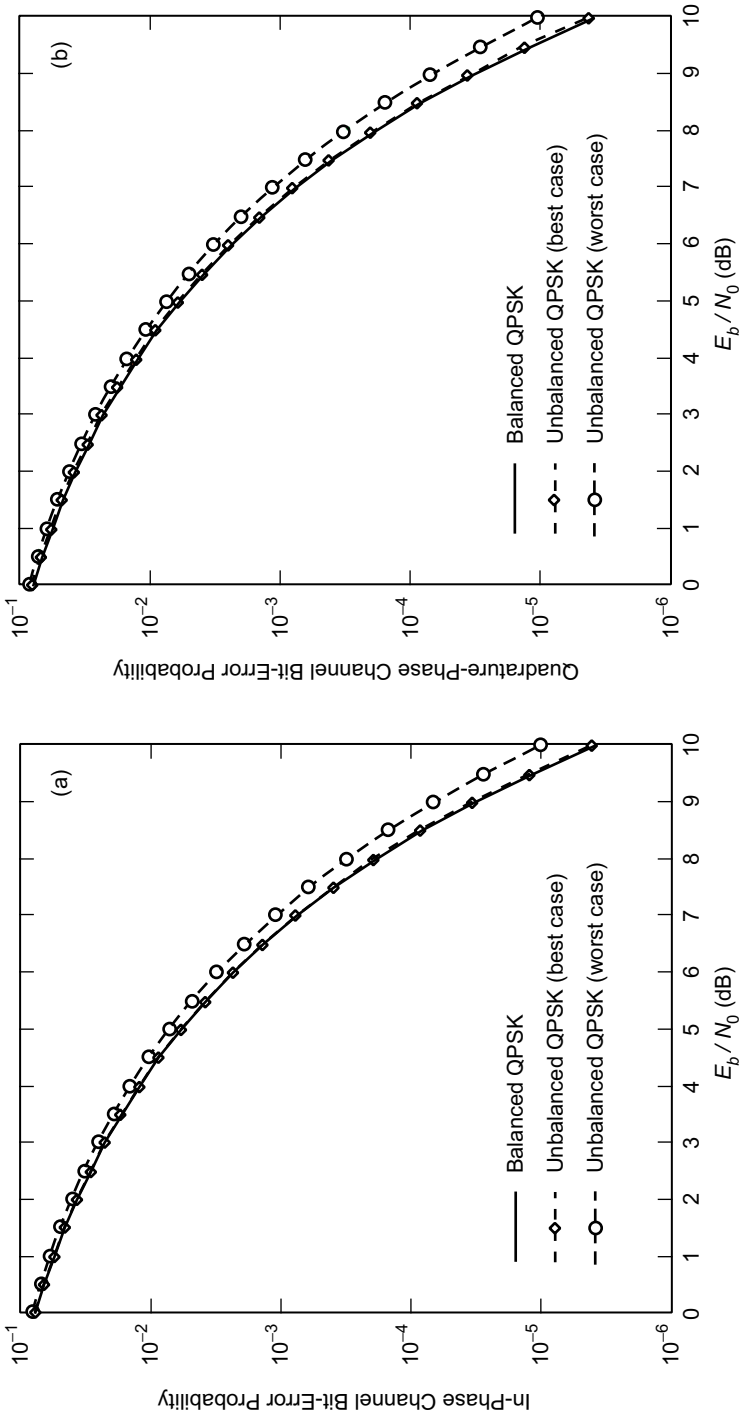

Fig. 2-4. Bit-error performance of imbalanced QPSK signals: (a) in-phase channel and (b) quadrature channel.

where the only changes in the parameters of (2.7-3) are that α_Q, β_Q, and γ_Q are now each multiplied by the I-Q amplitude imbalance parameter, Γ.

The carrier-tracking loop assumed in Ref. 9 is a slightly modified version of that used for QPSK, in which a half-symbol delay is added to its I arm so that the symbols on both arms are aligned in forming the $IQ\left(Q^2 - I^2\right)$ error signal. This loop as well as the optimum (based on maximum a posteriori (MAP) estimation) OQPSK loop, which exhibits only a 180-deg phase ambiguity, are discussed in Ref. 12. The evaluation of the steady-state lock point of the loop was considered in Ref. 9 and was determined numerically. The average BEP is still determined from (2.7-6) (again assuming perfect 90-deg phase ambiguity resolution), but the conditional I and Q BEPs are now specified by

$$
\begin{aligned}
&P_{bI}\left(E;\phi\right) = \\
&\frac{1}{16}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\cos\left(\phi+\Delta\theta_c\right)+\sin\phi\right]\right) \\
&+\frac{1}{16}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\cos\left(\phi+\Delta\theta_c\right)-\Gamma\Gamma_Q\sin\left(\phi+\Delta\theta_{cQ}\right)\right]\right) \\
&+\frac{1}{16}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma_I\cos\left(\phi+\Delta\theta_{c_I}+\Delta\theta_c\right)-\Gamma\sin\phi\right]\right) \\
&+\frac{1}{16}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma_I\cos\left(\phi+\Delta\theta_{c_I}+\Delta\theta_c\right)+\Gamma\Gamma_Q\sin\left(\phi+\Delta\theta_{cQ}\right)\right]\right) \\
&+\frac{1}{8}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\cos\left(\phi+\Delta\theta_c\right)-\frac{\Gamma\Gamma_Q}{2}\sin\left(\phi+\Delta\theta_{cQ}\right)+\frac{\Gamma}{2}\sin\phi\right]\right) \\
&+\frac{1}{8}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma_I\cos\left(\phi+\Delta\theta_{cI}+\Delta\theta_c\right)+\frac{\Gamma\Gamma_Q}{2}\sin\left(\phi+\Delta\theta_{cQ}\right)-\frac{\Gamma}{2}\sin\phi\right]\right)
\end{aligned}
\tag{2.7-11a}
$$

and

$$P_{bQ}(E;\phi) =$$

$$\frac{1}{16}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma\cos\phi - \sin(\phi+\Delta\theta_c)\right]\right)$$

$$+\frac{1}{16}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma\cos\phi + \Gamma_I\sin(\phi+\Delta\theta_{cI}+\Delta\theta_c)\right]\right)$$

$$+\frac{1}{16}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma\Gamma_Q\cos(\phi+\Delta\theta_{cQ}) + \sin(\phi+\Delta\theta_c)\right]\right)$$

$$+\frac{1}{16}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma\Gamma_Q\cos(\phi+\Delta\theta_{cQ}) - \Gamma_I\sin(\phi+\Delta\theta_{cI}+\Delta\theta_c)\right]\right)$$

$$+\frac{1}{8}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma\cos\phi + \frac{\Gamma_I}{2}\sin(\phi+\Delta\theta_{cI}+\Delta\theta_c) + \frac{1}{2}\sin(\phi+\Delta\theta_c)\right]\right)$$

$$+\frac{1}{8}\operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\left[\Gamma\Gamma_Q\cos(\phi+\Delta\theta_{cQ}) - \frac{\Gamma_I}{2}\sin(\phi+\Delta\theta_{cI}+\Delta\theta_c)\right.\right.$$

$$\left.\left.+\frac{1}{2}\sin(\phi+\Delta\theta_c)\right]\right) \qquad \text{(2.7-11b)}$$

Substituting (2.7-7) together with (2.7-11a) and (2.7-11b) in (2.7-6) gives the desired average BEP of the I and Q channels for any degree of modulator imbalance. Note again that, in general, the error probability performances of the I and Q channels are not identical.

For the same maximum amplitude imbalance, maximum phase imbalance, and maximum I-Q quadrature imbalances as for the QPSK case and in addition an I-Q amplitude imbalance (Γ) of -0.2 dB (corresponding to an actual Q-channel power that is 0.4 dB less than that in the I channel), Figs. 2-5(a) and 2-5(b) plot the I and Q average BEPs as computed from (2.7-6) for the best and worst combinations of imbalance conditions. These results also include the effect of a finite loop SNR of the ϕ process, $\rho_\phi = \rho_{4\phi}/16$, which was chosen equal to 22 dB and held constant along the curves. The case of perfectly balanced QPSK is included in these plots for comparison purposes. The curve labeled

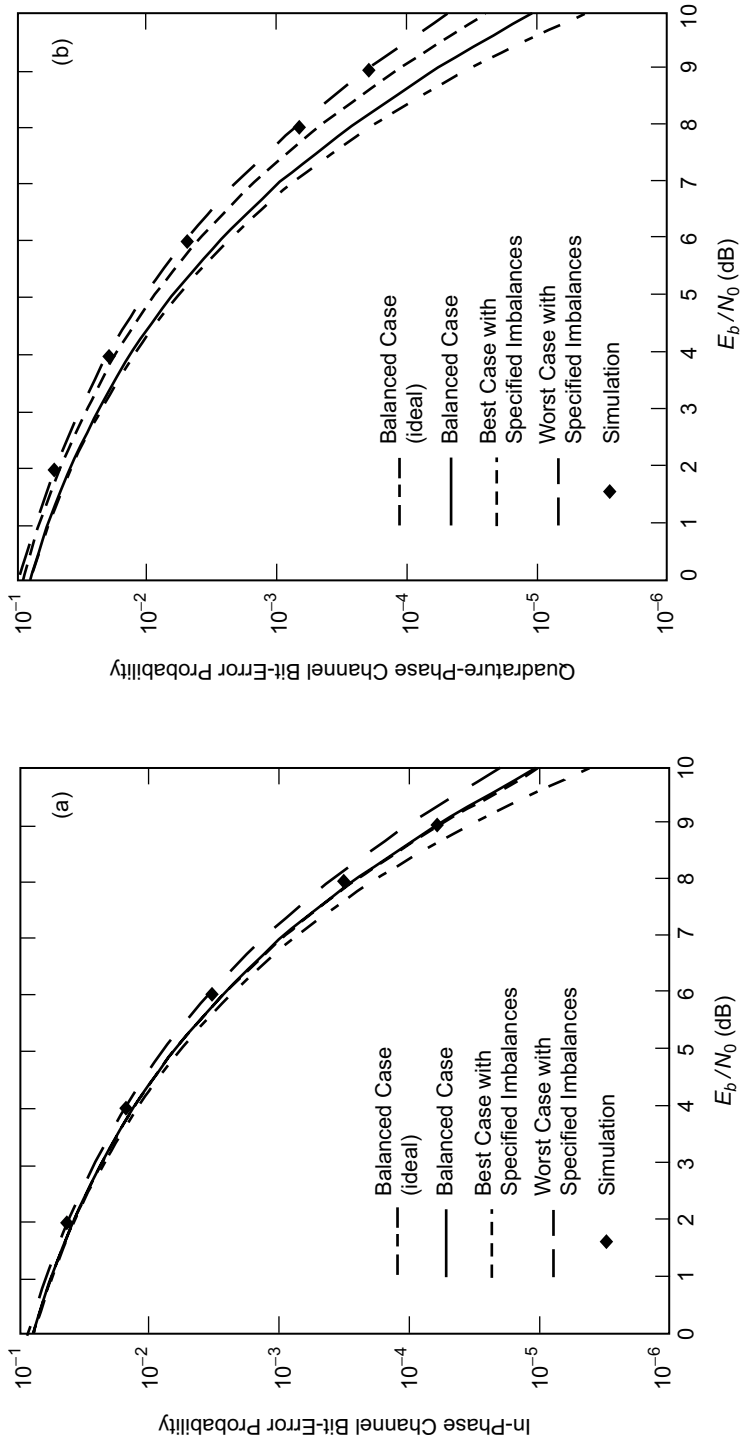

Fig. 2-5. Bit-error performance of OQPSK signals under imperfect carrier synchronization: (a) in-phase channel and (b) quadrature-phase channel.

balanced QPSK (ideal) refers to the case where the loop SNR is assumed infinite, as was the case shown in Figs. 2-4(a) and 2-4(b). Finally, simulation points that agree with the analytical results are also included in Figs. 2-5(a) and 2-5(b). We observe from these figures that the worst imbalance condition results in an E_b/N_0 loss of 0.61 dB for the I channel and 1.08 dB for the Q-channel at an average BEP of 10^{-4}, the larger loss for the Q channel coming as a result of its 0.4-dB power deficiency caused by the I-Q amplitude imbalance. When the I and Q results are averaged, the overall E_b/N_0 degradation becomes 0.86 dB. If perfect carrier synchronization had been assumed, then as shown in Ref. 9, these worst-case losses would be reduced to 0.34 dB for the I channel and 0.75 dB for the Q channel, which translates to a 0.58-dB average performance degradation.

Aside from intrachannel and interchannel amplitude and phase imbalances, the inclusion of a fully saturated RF amplifier modeled by a bandpass hard limiter in the analytical model causes additional degradation in system performance. The performance of OQPSK on such a nonlinear channel was studied in Ref. 10, using the same modulator imbalance model as previously discussed above. The results are summarized as follows.

The transmitter is the same as that illustrated in Fig. 2-3 (with the inclusion of the half-symbol delay in the Q channel as previously discussed), the output of which is now passed through a nonlinear amplifier composed of the cascade of a hard limiter and a bandpass filter (a bandpass hard limiter [13]). The hard limiter clips its input signal at levels $\pm\sqrt{2P_1}\,(\pi/4)$, and the bandpass (zonal) filter removes all the harmonics except for the one at the carrier frequency. The resulting bandpass hard-limited OQPSK signal is a constant envelope signal that has the form

$$\hat{s}\,(t) = \sqrt{2P_1}\cos\left(2\pi f_c t + \theta_d\,(t)\right) \tag{2.7-12}$$

where $P_1 = P\left(\beta_I^2 + \gamma_I^2\right)$ with β_I, γ_I as defined in (2.7-3) and[8]

$$\theta_d\,(t) = \tan^{-1}\frac{\gamma_I}{\beta_I} - \tan^{-1}\left(\frac{Gm_Q\left(t - \dfrac{T_s}{2}\right)\cos\Delta\theta + A\cos\psi}{m_I\,(t) + Gm_Q\left(t - \dfrac{T_s}{2}\right)\sin\Delta\theta + A\sin\psi}\right) \tag{2.7-13}$$

with

[8] The arctangents in (2.7-13) are taken in their principal value sense. Thus, adding π to some of these values is required to place $\theta_d\,(t)$ into its appropriate quadrant.

$$\left.\begin{aligned} G &= \sqrt{\frac{\beta_Q^2+\gamma_Q^2}{\beta_I^2+\gamma_I^2}} \\ A &= \sqrt{\frac{(\alpha_I-\gamma_Q)^2+(\alpha_Q+\delta_I)^2}{\beta_I^2+\gamma_I^2}} \\ \Delta\theta &= \tan^{-1}\frac{\gamma_Q}{\beta_Q}-\tan^{-1}\frac{\gamma_I}{\beta_I} \\ \psi &= \tan^{-1}\frac{\alpha_I-\gamma_Q}{\alpha_Q+\delta_I}-\tan^{-1}\frac{\gamma_I}{\beta_I} \end{aligned}\right\} \tag{2.7-14}$$

Since in any half symbol interval, $m_I(t)$ and $m_Q(t-[T_s/2])$ only take on values ± 1, then in that same interval, $\theta_d(t)$ takes on only one of four equiprobable values, namely, $\theta_{1,1}, \theta_{-1,1}, \theta_{1,-1}, \theta_{-1,-1}$, where the subscripts correspond, respectively, to the values of the above two modulations.

The average BEP is again computed from (2.7-6) together with (2.7-7), where the conditional BEPs are now given by [10, Eqs. (10a) and (10b)]

$$P_{bI}(E;\phi) = \overline{\frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{2E_b'}{N_0}}\cos\left(\frac{\theta_d^{(1)}-\theta_d^{(2)}}{2}\right)\left|\cos\left(\frac{\theta_d^{(1)}+\theta_d^{(2)}}{2}+\phi\right)\right|\right)} \tag{2.7-15}$$

$$P_{bQ}(E;\phi) = \overline{\frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{2E_b'}{N_0}}\cos\left(\frac{\theta_d^{(2)}-\theta_d^{(3)}}{2}\right)\left|\cos\left(\frac{\theta_d^{(2)}+\theta_d^{(3)}}{2}+\phi\right)\right|\right)}$$

where $\theta_d^{(j)}$ is the value of the symbol phase $\theta_d(t)$ in the interval $(j-1)T_s/2 \le t \le jT_s/2$, the overbar denotes the statistical average over these symbol phases, and $E_b' = P_1 T/2T_s = \left(\beta_I^2+\gamma_I^2\right)PT_s/2 = \left(\beta_I^2+\gamma_I^2\right)E_b$ is the actual I-channel bit energy. Using now the steady-state lock point (irreducible carrier phase error) found numerically in Ref. 10 for this scenario, the average overall and I and Q BEPs are illustrated in Figs. 2-6(a), 2-6(b), and 2-6(c) using parameters identical to those used in arriving at Figs. 2-5(a) and 2-5(b). The final result is that, in the presence of modulator imbalance, the nonlinear amplifier tends to produce a more balanced signal constellation, and thus, the relative BEP performance

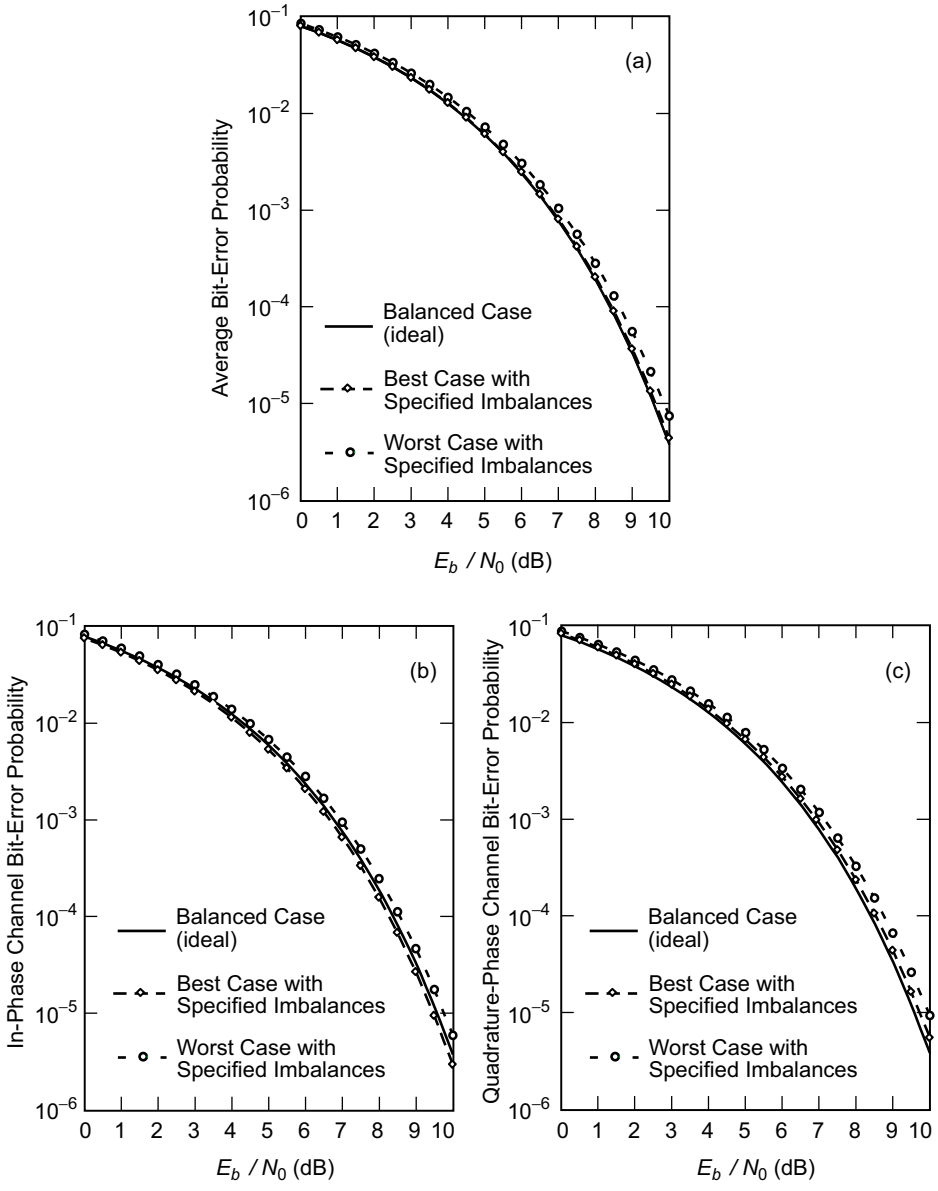

Fig. 2-6. Bit-error performance of nonlinear OQPSK links with imperfect carrier synchronization (i.e., with a carrier-tracking loop SNR fixed at 22 dB): (a) overall channel, (b) in-phase channel, and (c) quadrature-phase channel.

between the I and Q channels is itself more balanced. Furthermore, the average BEPs themselves are much closer to that of a perfectly balanced OQPSK system than those found for the linear channel.

2.7.2 Data Imbalance

The presence of data imbalance (positive and negative bits have different a priori probabilities of occurrence) in the transmitted waveform results in the addition of a discrete spectral component at dc to the continuous PSD component described by (2.5-1). Specifically, if p denotes the probability of a mark (+1), then the total PSD is given by [11, Eq. (1-19)]

$$S(f) = PT_s \left[\frac{1}{T_s} (1-2p)^2 \delta(f) + 4p(1-p) \frac{\sin^2 \pi f T_s}{(\pi f T_s)^2} \right] \tag{2.7-16}$$

Clearly, for the balanced data case, i.e., $p = 1/2$, (2.7-16) reduces to (2.5-1). Since the total power in the transmitted signal is now split between an unmodulated tone at the carrier frequency and a data-bearing component, the carrier tracking process at the receiver (which is designed to act only on the latter) becomes affected even with perfect modulator balance. The degrading effects of a residual carrier on the Costas loop performance for binary PSK are discussed in Ref. 14. The extension to QPSK and OQPSK modulations is straightforward and not pursued here.

Further on in this monograph in our discussion of simulation models and performance, we shall talk about various types of filtered QPSK (which would then no longer be constant envelope). At that time, we shall observe that the combination of data imbalance and filtering produces additional discrete spectral harmonics occurring at integer multiples of the symbol rate.

2.8 Continuous Phase Modulation

Continuing with our discussion of strictly constant envelope modulations, we now turn our attention to the class of schemes referred to as continuous phase frequency modulation (CPFM) or more simply continuous phase modulation (CPM). The properties and performance (bandwidth/power) characteristics of this class of modulations are sufficiently voluminous to fill a textbook of their own [15]. Thus, for the sake of brevity, we shall only investigate certain special cases of CPM that have gained popularity in the literature and have also been put to practice.

CPM schemes are classified as being full response or partial response, depending, respectively, on whether the modulating frequency pulse is of a single bit duration or longer. Within the class of full response CPMs, the subclass of schemes having modulation index 0.5 but arbitrary frequency pulse shape results in a form of generalized MSK [16].[9] Included as popular special cases are MSK, originally invented by Doelz and Heald, as disclosed in a 1961 U.S. patent [19], having a rectangular frequency pulse shape, and Amoroso's sinusoidal frequency-shift-keying (SFSK) [20], possessing a sinusoidal (raised cosine) frequency pulse shape. The subclass of full-response schemes with rectangular frequency pulse but arbitrary modulation index is referred to as continuous phase frequency-shift-keying (CPFSK) [21], which, for all practical purposes, served as the precursor to what later became known as CPM itself. Within the class of partial-response CPMs, undoubtedly the most popular scheme is that of Gaussian minimum-shift-keying (GMSK) which, because of its excellent bandwidth efficiency, has been adopted as a European standard for personal communication systems (PCSs). In simple terms, GMSK is a partial-response CPM scheme obtained by filtering the rectangular frequency pulses characteristic of MSK with a filter having a Gaussian impulse response prior to frequency modulation of the carrier.

In view of the above considerations, in what follows, we shall focus our CPM discussion only on MSK, SFSK, and GMSK, in each case presenting results for their spectral and power efficiency behaviors. Various representations of the transmitter, including the all-important equivalent I-Q one, will be discussed as well as receiver performance, both for ideal and nonideal (modulator imbalance) conditions.

2.8.1 Full Response—MSK and SFSK

While the primary intent of this section of the monograph is to focus specifically on the properties and performance of MSK and SFSK in the form they are most commonly known, the reader should bear in mind that many of these very same characteristics, e.g., transmitter/receiver implementations, equivalent I-Q signal representations, spectral and error probability analysis, apply equally well to generalized MSK. Whenever convenient, we shall draw attention to these analogies so as to alert the reader to the generality of our discussions. We begin the mathematical treatment by portraying MSK as a special case of the more general CPM signal, whose characterization is given in the next section.

---

[9] Several other authors [17,18] coined the phrase "generalized MSK" to represent generalizations of MSK other than by pulse shaping.

2.8.1.1 Continuous Phase Frequency Modulation Representation. A binary single-mode (one modulation index for all transmission intervals) CPM signal is a constant envelope waveform that has the generic form (see the implementation in Fig. 2-7)

$$s(t) = \sqrt{\frac{2E_b}{T_b}} \cos\left(2\pi f_c t + \phi\left(t, \boldsymbol{\alpha}\right) + \phi_0\right), \quad nT_b \le t \le (n+1)T_b \tag{2.8-1}$$

where, as before, E_b and T_b respectively denote the energy and duration of a bit ($P = E_b/T_b$ is the signal power), and f_c is the carrier frequency. In addition, $\phi\left(t, \boldsymbol{\alpha}\right)$ is the phase modulation process that is expressable in the form

$$\phi\left(t, \boldsymbol{\alpha}\right) = 2\pi \sum_{i \le n} \alpha_i h q\left(t - iT_b\right) \tag{2.8-2}$$

where $\boldsymbol{\alpha} = (\cdots, \alpha_{-2}, \alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \cdots)$ is an independent, identically distributed (i.i.d.) binary data sequence, with each element taking on equiprobable values ± 1, $h = 2\Delta f T_b$ is the modulation index (Δf is the peak frequency deviation of the carrier), and $q(t)$ is the normalized phase-smoothing response that defines how the underlying phase, $2\pi\alpha_i h$, evolves with time during the associated bit interval. Without loss of generality, the arbitrary phase constant, ϕ_0, can be set to zero.

For our discussion here it is convenient to identify the derivative of $q(t)$, namely,

$$g(t) = \frac{dq(t)}{dt} \tag{2.8-3}$$

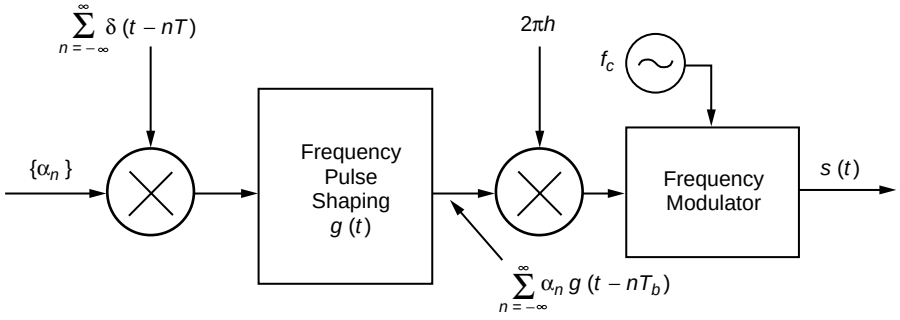

Fig. 2-7. CPM transmitter.

which represents the instantaneous frequency pulse (relative to the nominal carrier frequency, f_c) in the zeroth signaling interval. In view of (2.8-3), the phase smoothing response is given by

$$q(t) = \int_{-\infty}^{t} g(\tau)\, d\tau \tag{2.8-4}$$

which, in general, extends over infinite time. For full response CPM schemes, as will be the case of interest here, $q(t)$ satisfies the following:

$$q(t) = \begin{cases} 0, & t \leq 0 \\ \dfrac{1}{2}, & t \geq T_b \end{cases} \tag{2.8-5}$$

and, thus, the frequency pulse, $g(t)$, is nonzero only over the bit interval, $0 \leq t \leq T_b$. In view of (2.8-5), we see that the ith data symbol, α_i, contributes a phase change of $\pi\alpha_i h$ rad to the total phase for all time after T_b seconds of its introduction, and, therefore, this fixed phase contribution extends over all future symbol intervals. Because of this overlap of the phase smoothing responses, the total phase in any signaling interval is a function of the present data symbol as well as all of the past symbols, and accounts for the memory associated with this form of modulation. Consequently, in general, optimum detection of CPM schemes must be performed by a maximum-likelihood sequence estimator (MLSE) form of receiver [1] as opposed to bit-by-bit detection, which is optimum for memoryless modulations such as conventional binary FSK with discontinuous phase.

As previously mentioned, MSK is a full-response CPM scheme with a modulation index $h = 0.5$ and a rectangular frequency pulse mathematically described by

$$g(t) = \begin{cases} \dfrac{1}{2T_b}, & 0 \leq t \leq T_b \\ 0, & \text{otherwise} \end{cases} \tag{2.8-6}$$

For SFSK, one of the generalized MSK schemes mentioned in the introduction, $g(t)$, would be a raised cosine pulse given by

$$g(t) = \begin{cases} \dfrac{1}{2T_b}\left[1 - \cos\left(\dfrac{2\pi t}{T_b}\right)\right], & 0 \leq t \leq T_b \\ 0, & \text{otherwise} \end{cases} \tag{2.8-7}$$

The associated phase pulses defined by (2.8-4) are

$$q(t) = \begin{cases} \dfrac{t}{2T_b}, & 0 \le t \le T_b \\ \dfrac{1}{2}, & t \ge T_b \end{cases} \tag{2.8-8}$$

for MSK and

$$q(t) = \begin{cases} \dfrac{1}{2T_b}\left[t - \dfrac{\sin 2\pi t/T_b}{2\pi/T_b}\right], & 0 \le t \le T_b \\ \dfrac{1}{2}, & t \ge T_b \end{cases} \tag{2.8-9}$$

for SFSK.

Finally, substituting $h = 0.5$ and $g(t)$ of (2.8-6) in (2.8-1) combined with (2.8-2) gives the CPM representations of MSK and SFSK, respectively, as

$$s_{\mathrm{MSK}}(t) = \sqrt{\frac{2E_b}{T_b}} \cos\left(2\pi f_c t + \frac{\pi}{2T_b}\sum_{i\le n}\alpha_i\left(t - iT_b\right)\right), \quad nT_b \le t \le (n+1)T_b \tag{2.8-10}$$

and

$$s_{\mathrm{SFSK}}(t) = \sqrt{\frac{2E_b}{T_b}} \cos\left(2\pi f_c t + \frac{\pi}{2T_b}\sum_{i\le n}\alpha_i\left[t - iT_b - \frac{\sin 2\pi\left(t - iT_b\right)/T_b}{2\pi/T_b}\right]\right), \quad nT_b \le t \le (n+1)T_b \tag{2.8-11}$$

both of which are implemented as in Fig. 2-7, using $g(t)$ of (2.8-6) or (2.8-7) as appropriate.

Associated with MSK (or SFSK) is a phase trellis that illustrates the evolution of the phase process with time, corresponding to all possible transmitted sequences. For MSK, the phase variation with time is linear [see (2.8-10)], and, thus, paths in the phase trellis are straight lines with a slope of $\pm\pi/2T_b$. Figure 2-8 is an illustration of the MSK phase trellis where the branches are labeled with the data bits that produce the corresponding phase transition. Note that

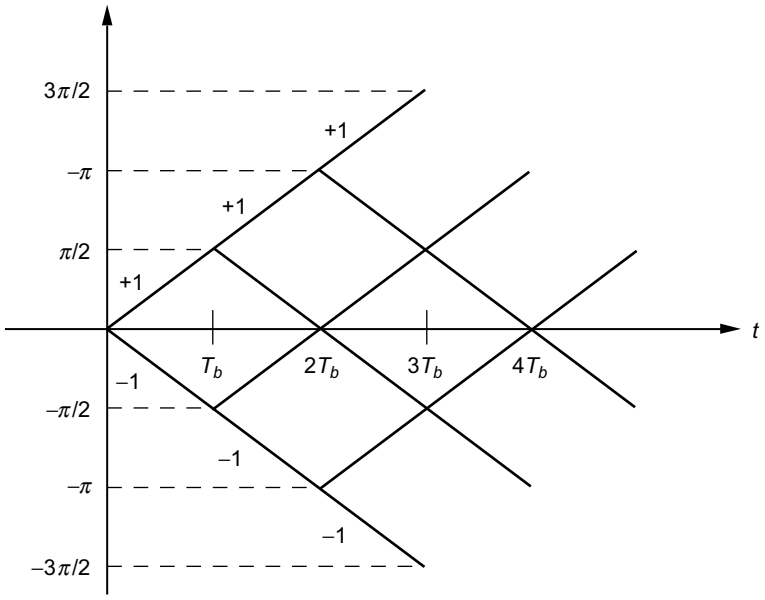

Fig. 2-8. Phase trellis (time-varying) for conventional MSK. Phase states (mod 2π) are $(0, \pi)$ for *n* even and $(\pi/2, 3\pi/2)$ for *n* odd.

the change in phase over a single bit time is either $\pi/2$ or $-\pi/2$, depending on the polarity of the data bit, α_i, corresponding to that bit time. Also note that the trellis is time-varying in that the phase states (modulo 2π) alternate between 0 and π at even multiples of the bit time and $\pi/2$ and $3\pi/2$ at odd multiples of the bit time. For SFSK, the phase trellis would appear as in Fig. 2-8 with, however, a sinusoidal variation in phase superimposed over the straight line paths. Here again the change in phase over a single bit time would be either $\pi/2$ or $-\pi/2$, depending on the polarity of the data bit, α_i, corresponding to that bit time.

2.8.1.2 Equivalent I-Q Representation of MSK. Although, as stated above, CPM schemes, because of their inherent memory, require a memory-type of detection, e.g., MLSE, full-response modulations with $h = 0.5$ such as MSK and SFSK can in fact be detected using a memoryless I-Q form of receiver. The reason for this is that for these modulations the transmitter can be implemented in an I-Q form analogous to that of OQPSK. To see this mathematically, we first rewrite the excess phase in the nth transmission interval of the MSK signal in (2.8-10) as

$$\phi(t,\boldsymbol{\alpha}) = \frac{\pi}{2T_b}\sum_{i\le n}\alpha_i(t-iT_b) = \alpha_n\frac{\pi}{2T_b}(t-nT_b) + \frac{\pi}{2}\sum_{i\le n-1}\alpha_i = \alpha_n\frac{\pi}{2T_b}t + x_n,$$

$$nT_b \le t \le (n+1)T_b \qquad (2.8\text{-}12)$$

where $(\pi/2)\sum_{i\le n-1}\alpha_i$ is the accumulated phase at the beginning of the nth transmission interval that is equal to an odd integer (positive or negative) multiple of $\pi/2$ when n is odd and an even integer (positive or negative) multiple of $\pi/2$ when n is even, and x_n is a phase constant required to keep the phase continuous at the data transition points $t = nT_b$ and $t = (n+1)T_b$. Note also that x_n represents the y-intercept (when reduced modulo 2π) of the path in the phase trellis that represents $\phi(t,\boldsymbol{\alpha})$. In the previous transmission interval, the excess phase is given by

$$\phi(t,\boldsymbol{\alpha}) = \alpha_n\frac{\pi}{2T_b}\big(t-(n-1)T_b\big) + \frac{\pi}{2}\sum_{i\le n-2}\alpha_i = \alpha_{n-1}\frac{\pi}{2T_b}t + x_{n-1},$$

$$(n-1)T_b \le t \le nT_b \qquad (2.8\text{-}13)$$

For phase continuity at $t = nT_b$, we require that

$$\alpha_n\frac{\pi}{2T_b}(nT_b) + x_n = \alpha_{n-1}\frac{\pi}{2T_b}(nT_b) + x_{n-1} \qquad (2.8\text{-}14)$$

or equivalently

$$x_n = x_{n-1} + \frac{\pi n}{2}(\alpha_{n-1} - \alpha_n) \qquad (2.8\text{-}15)$$

Equation (2.8-15) is a recursive relation that allows x_n to be determined in any transmission interval given an initial condition, x_0.

We observe that $(\alpha_{n-1} - \alpha_n)/2$ is a ternary random variable (RV) taking on values $0, +1, -1$, with probabilities $1/2, 1/4, 1/4$, respectively. Therefore, from (2.8-15), when $\alpha_{n-1} = \alpha_n$, $x_n = x_{n-1}$, whereas when $\alpha_{n-1} \ne \alpha_n$, $x_n = x_{n-1} \pm \pi n$. If we arbitrarily choose the initial condition $x_0 = 0$, then we see that x_n takes on values of 0 or π (when reduced modulo 2π). Using this fact in (2.8-12) and applying simple trigonometry to (2.8-10), we obtain

$$s_{\mathrm{MSK}}(t) = \sqrt{\frac{2E_b}{T_b}}\left[\cos\phi\,(t,\alpha)\cos 2\pi f_c t - \sin\phi\,(t,\alpha)\sin 2\pi f_c t\right],$$

$$nT_b \le t \le (n+1)T_b \qquad (2.8\text{-}16)$$

where

$$\left.\begin{aligned}
\cos\phi\,(t,\alpha) &= \cos\left(\alpha_n \frac{\pi}{2T_b}t + x_n\right) = a_n \cos\frac{\pi}{2T_b}t, \quad a_n = \cos x_n = \pm 1 \\
\sin\phi\,(t,\alpha) &= \sin\left(\alpha_n \frac{\pi}{2T_b}t + x_n\right) = \alpha_n a_n \sin\frac{\pi}{2T_b}t = b_n \sin\frac{\pi}{2T_b}t, \\
b_n &= \alpha_n \cos x_n = \pm 1
\end{aligned}\right\} \qquad (2.8\text{-}17)$$

Finally, substituting (2.8-17) in (2.8-16) gives the I-Q representation of MSK as

$$s_{\mathrm{MSK}}(t) = \sqrt{\frac{2E_b}{T_b}}\left[a_n C\,(t)\cos 2\pi f_c t - b_n S\,(t)\sin 2\pi f_c t\right], \quad nT_b \le t \le (n+1)T_b \qquad (2.8\text{-}18)$$

where

$$\left.\begin{aligned}
C\,(t) &= \cos\frac{\pi t}{2T_b} \\
S\,(t) &= \sin\frac{\pi t}{2T_b}
\end{aligned}\right\} \qquad (2.8\text{-}19)$$

are the effective I and Q pulse shapes, and $\{a_n\}, \{b_n\}$, as defined in (2.8-17), are the effective I and Q binary data sequences.

For SFSK, the representation of (2.8-18) would still be valid with a_n, b_n as defined in (2.8-17), but now the effective I and Q pulse shapes become

$$\left.\begin{aligned}
C\,(t) &= \cos\left[\frac{\pi}{2T_b}\left(t - \frac{\sin 2\pi t/T_b}{2\pi/T_b}\right)\right] \\
S\,(t) &= \sin\left[\frac{\pi}{2T_b}\left(t - \frac{\sin 2\pi t/T_b}{2\pi/T_b}\right)\right]
\end{aligned}\right\} \qquad (2.8\text{-}20)$$

To tie the representation of (2.8-18) back to that of FSK, we observe that

$$\left.\begin{aligned} C(t)\cos 2\pi f_c t &= \frac{1}{2}\cos\left[2\pi\left(f_c+\frac{1}{4T_b}\right)t\right] \\ &\quad+\frac{1}{2}\cos\left[2\pi\left(f_c-\frac{1}{4T_b}\right)t\right] \\ S(t)\sin 2\pi f_c t &= -\frac{1}{2}\cos\left[2\pi\left(f_c+\frac{1}{4T_b}\right)t\right] \\ &\quad+\frac{1}{2}\cos\left[2\pi\left(f_c-\frac{1}{4T_b}\right)t\right] \end{aligned}\right\} \tag{2.8-21}$$

Substituting (2.8-21) in (2.8-18) gives

$$s_{\mathrm{MSK}}(t)=\sqrt{\frac{2E_b}{T_b}}\left[\left(\frac{a_n+b_n}{2}\right)\cos\left[2\pi\left(f_c+\frac{1}{4T_b}\right)t\right]+\left(\frac{a_n-b_n}{2}\right)\cos\left[2\pi\left(f_c-\frac{1}{4T_b}\right)t\right]\right],\quad nT_b\le t\le(n+1)T_b \tag{2.8-22}$$

Thus, when $a_n=b_n\,(\alpha_n=1)$, we have

$$s_{\mathrm{MSK}}(t)=\sqrt{\frac{2E_b}{T_b}}\cos\left[2\pi\left(f_c+\frac{1}{4T_b}\right)t\right] \tag{2.8-23}$$

whereas when $a_n\neq b_n\,(\alpha_n=-1)$ we have

$$s_{\mathrm{MSK}}(t)=\sqrt{\frac{2E_b}{T_b}}\cos\left[2\pi\left(f_c-\frac{1}{4T_b}\right)t\right] \tag{2.8-24}$$

which establishes the desired connection.

Note from (2.8-19), that since $C(t)$ and $S(t)$ are offset from each other by a time shift of T_b seconds, it might appear that $s_{\text{MSK}}(t)$ of (2.8-18) is in the form of OQPSK with half-sinusoidal pulse shaping.[10] To justify that this is indeed the case, we must examine more carefully the effective I and Q data sequences $\{a_n\}, \{b_n\}$ in so far as their relationship to the input data sequence $\{\alpha_i\}$ and the rate at which they can change. Since the input α_n data bit can change every bit time, it might appear that the effective I and Q data bits, a_n and b_n, can also change every bit time. To the contrary, it can be shown that as a result of the phase continuity constraint of (2.8-15), $a_n = \cos x_n$ can change only at the zero crossings of $C(t)$, whereas $b_n = \alpha_n \cos x_n$ can change only at the zero crossings of $S(t)$. Since the zero crossings of $C(t)$ and $S(t)$ are each spaced $2T_b$ seconds apart, then a_n and b_n are constant over $2T_b$-second intervals (see Fig. 2-9 for an illustrative example). Further noting that the continuous waveforms $C(t)$ and $S(t)$ alternate in sign every $2T_b$ seconds, we can incorporate this sign change into the I and Q data sequences themselves and deal with a fixed, positive, time-limited pulse shape on each of the I and Q channels. Specifically, defining the pulse shape

$$p(t) = \begin{cases} \sin \dfrac{\pi t}{2T_b}, & 0 \le t \le 2T_b \\ 0, & \text{otherwise} \end{cases} \tag{2.8-25}$$

then the I-Q representation of MSK can be rewritten in the form

$$s_{\text{MSK}}(t) = \sqrt{\frac{2E_b}{T_b}}\left[d_c(t)\cos 2\pi f_c t - d_s(t)\sin 2\pi f_c t\right] \tag{2.8-26}$$

where

$$\left.\begin{aligned} d_c(t) &= \sum_n c_n p\big(t - (2n-1)T_b\big) \\ d_s(t) &= \sum_n d_n p\left(t - 2nT_b\right) \end{aligned}\right\} \tag{2.8-27}$$

with

[10] A similar statement can be made for SFSK, where the pulse shaping is now described by (2.8-20).

n	α_n	$x_n \pmod{2\pi}$	a_n	b_n	Time Interval
0	1	0	1	1	$0 \le t \le T_b$
1	–1	π	–1	1	$T_b \le t \le 2T_b$
2	–1	π	–1	1	$2T_b \le t \le 3T_b$
3	1	0	1	1	$3T_b \le t \le 4T_b$
4	1	0	1	1	$4T_b \le t \le 5T_b$
5	1	0	1	1	$5T_b \le t \le 6T_b$
6	–1	0	1	–1	$6T_b \le t \le 7T_b$
7	1	π	–1	–1	$7T_b \le t \le 8T_b$
8	–1	π	–1	1	$8T_b \le t \le 9T_b$

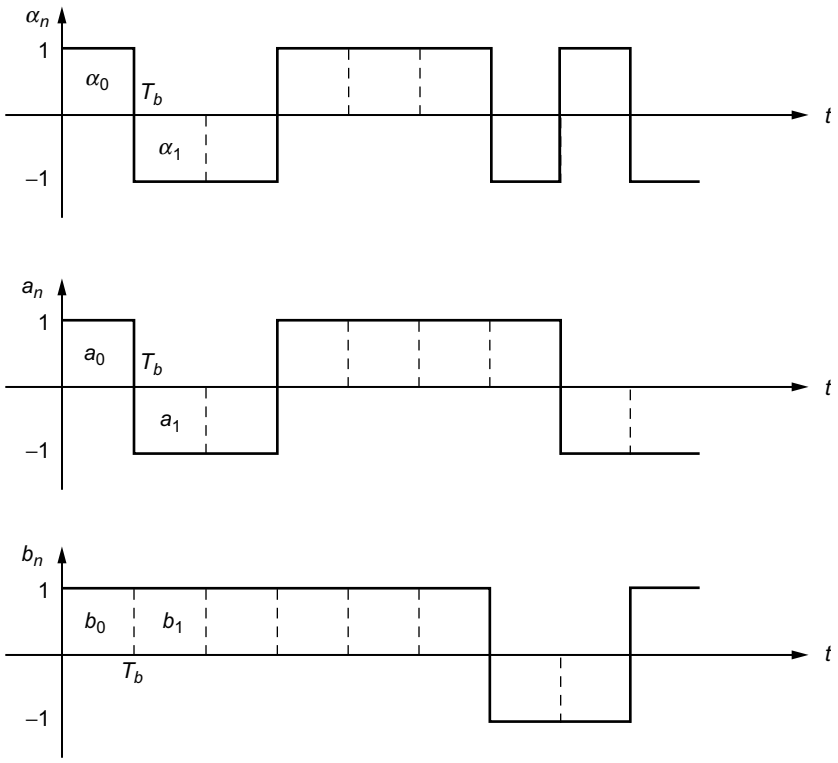

Fig. 2-9. An example of the equivalent I and Q data sequences represented as rectangular pulse streams. Redrawn from [1].

$$\left.\begin{aligned} c_n &= (-1)^n a_{2n-1} \\ d_n &= (-1)^n b_{2n} \end{aligned}\right\} \tag{2.8-28}$$

To complete the analogy between MSK and sinusoidally pulse shaped OQPSK, we must examine the manner in which the equivalent I and Q data sequences needed in (2.8-28) are obtained from the input data sequence $\{\alpha_n\}$. Without going into great mathematical detail, we can say that it can be shown that the sequences $\{a_{2n-1}\}$ and $\{b_{2n}\}$ are the odd/even split of a sequence, $\{v_n\}$, which is the differentially encoded version of $\{\alpha_n\}$, i.e., $v_n = \alpha_n v_{n-1}$ (see Fig. 2-10 for an illustrative example). Finally, the I-Q implementation of MSK as described by (2.8-26)–(2.8-28) is illustrated in Fig. 2-11. As anticipated, we observe that this figure resembles a transmitter for OQPSK except that here, the pulse shaping is half-sinusoidal (of symbol duration $T_s = 2T_b$) rather than rectangular; in addition, we see that a differential encoder is applied to the input data sequence prior to splitting it into even and odd sequences, each at a rate $1/T_s$. The interpretation of MSK as a special case of OQPSK with sinusoidal pulse shaping along with trade-offs and comparisons between the two modulations is further discussed in Refs. 22 and 23.

Before concluding this section, we note that the alternative representation of MSK as in (2.8-22) can be also expressed in terms of the differentially encoded bits, v_n. In particular,

For n odd

$$s_{\text{MSK}}(t) = \sqrt{\frac{2E_b}{T_b}}\left[\left(\frac{v_{n-1}+v_n}{2}\right)\cos\left[2\pi\left(f_c+\frac{1}{4T_b}\right)t\right]\right.$$
$$\left.-\left(\frac{v_{n-1}-v_n}{2}\right)\cos\left[2\pi\left(f_c-\frac{1}{4T_b}\right)t\right]\right],$$
$$nT_b \le t \le (n+1)T_b \tag{2.8-29a}$$

For n even

$$s_{\text{MSK}}(t) = \sqrt{\frac{2E_b}{T_b}}\left[\left(\frac{v_{n-1}+v_n}{2}\right)\cos\left[2\pi\left(f_c+\frac{1}{4T_b}\right)t\right]\right.$$
$$\left.+\left(\frac{v_{n-1}-v_n}{2}\right)\cos\left[2\pi\left(f_c-\frac{1}{4T_b}\right)t\right]\right],$$
$$nT_b \le t \le (n+1)T_b \tag{2.8-29b}$$

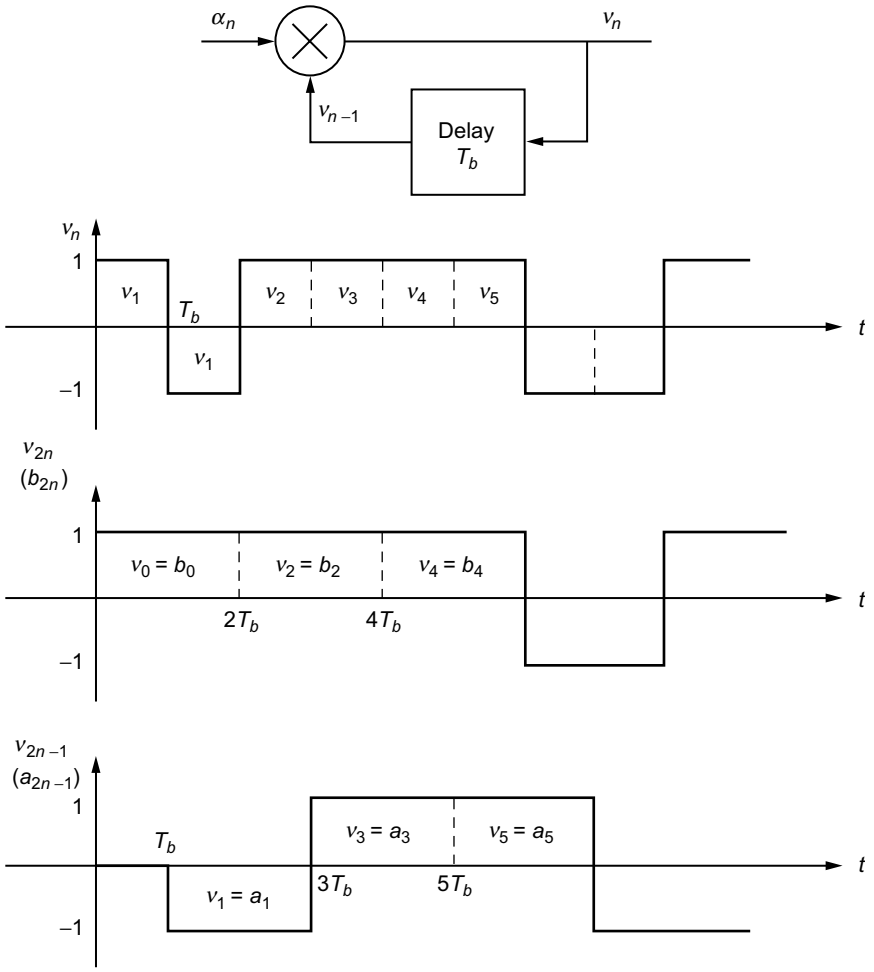

Fig. 2-10. An example of the equivalence between differentially encoded inputs bits and effective I and Q bits. Redrawn from [1].

Combining these two results we get

$$s_{\mathrm{MSK}}(t) = \sqrt{\frac{2E_b}{T_b}}\left[\left(\frac{v_{n-1}+v_n}{2}\right)\cos\left[2\pi\left(f_c+\frac{1}{4T_b}\right)t\right]\right.$$

$$\left.+(-1)^n\left(\frac{v_{n-1}-v_n}{2}\right)\cos\left[2\pi\left(f_c-\frac{1}{4T_b}\right)t\right]\right],$$

$$nT_b \le t \le (n+1)T_b \qquad (2.8\text{-}30)$$

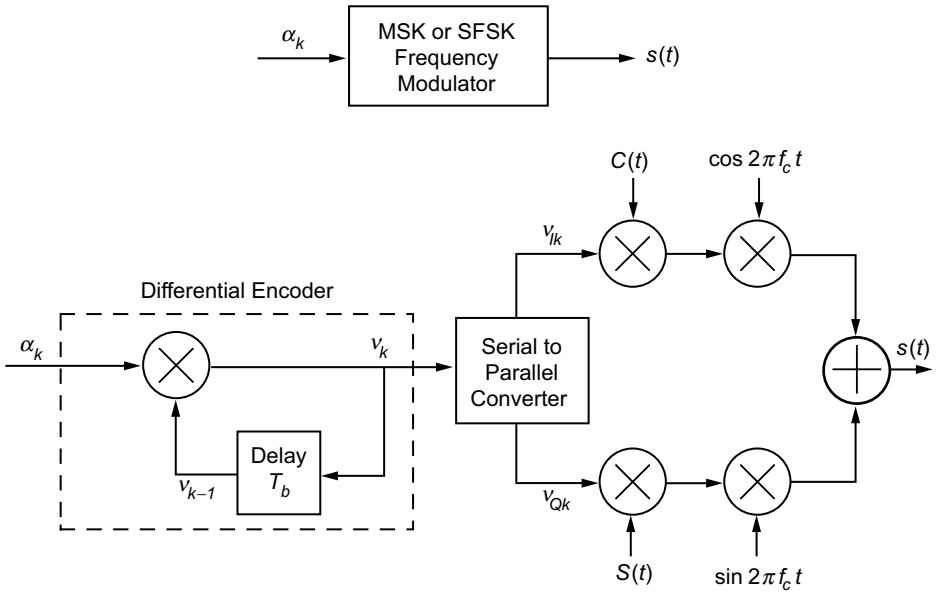

Fig. 2-11. CPM and equivalent I-Q implementations of MSK or SFSK.

2.8.1.3 Precoded MSK. The differential encoder that precedes the I-Q portion of the transmitter in Fig. 2-11 requires a compensating differential decoder at the receiver following I-Q demodulation and detection (see Fig. 2-12). Such a combination of differential encoding at the transmitter and differential decoding at the receiver results in a loss in power performance relative to that obtained by conventional OQPSK (this will be discussed in more detail later on in the chapter). It is possible to modify MSK to avoid such a loss by first recognizing that the CPM form of modulator in Fig. 2-7 for implementing MSK can be preceded by the cascade of a differential encoder and a differential decoder without affecting its output (Fig. 2-13). That is, the cascade of a differential encoder and a differential decoder produces unity transmission—input = output. Thus, comparing Fig. 2-13 with Fig. 2-11, we observe that precoding the CPM form of MSK modulator with a differential decoder, resulting in what is referred to as precoded MSK [1, Chap. 10] will be equivalent to the I-Q implementation of the latter without the differential encoder at its input (see Fig. 2-14), and thus the receiver for precoded MSK is that of Fig. 2-12 without the differential decoder at its output. A similar precoding applied to SFSK would also allow for dispensing with the differential decoder at the output of its I-Q receiver. Finally, we note that both MSK (or SFSK) and its precoded version have identical spectral characteristics and, consequently, for all practical purposes, the improvement in power performance provided by the latter comes at no expense.

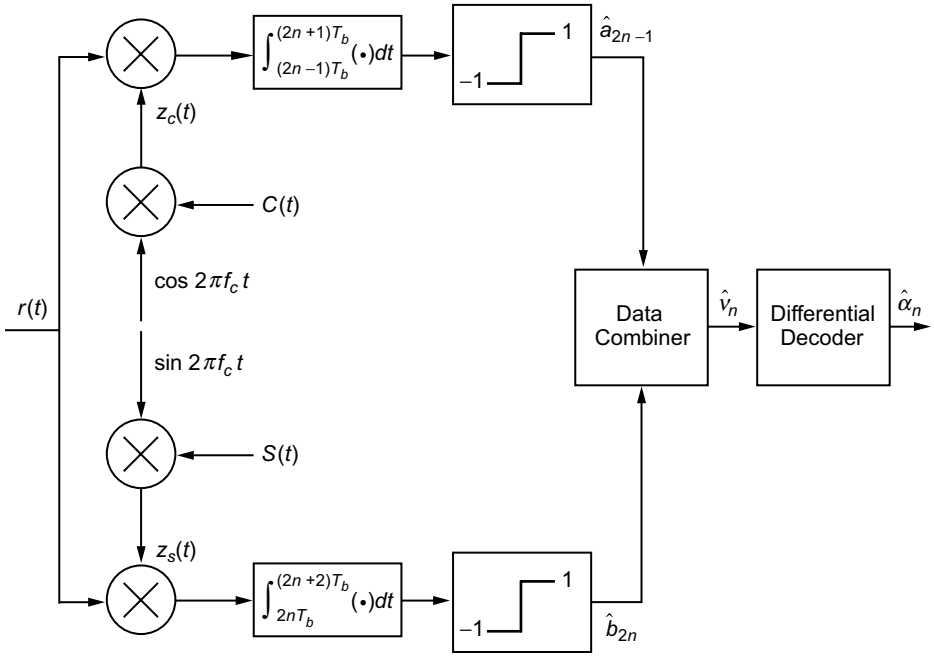

Fig. 2-12. An I-Q receiver implementation of MSK.

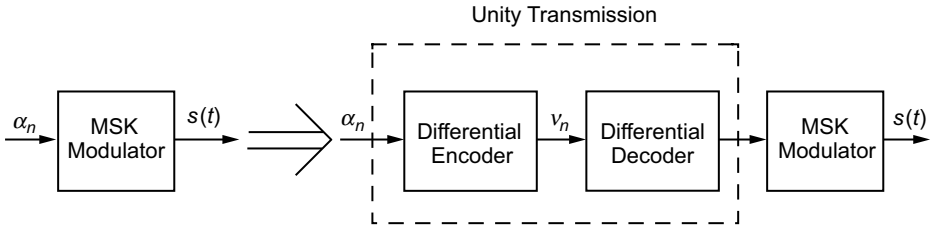

Fig. 2-13. Two equivalent MSK transmitters.

2.8.1.4 Spectral Characteristics. The ability to express MSK in the offset I-Q form of (2.8-18) allows for simple evaluation of its PSD. In particular, for a generic offset I-Q modulation formed by impressing two lowpass modulations (random pulse trains of rate $1/2T_b$) of equal power and pulse shape on inphase and quadrature carriers, i.e.,

$$s(t) = Am_I(t)\cos 2\pi f_c t - Am_Q(t)\sin 2\pi f_c t,$$

$$m_I(t) = \sum_n a_n p(t - 2nT_b), \quad m_Q(t) = \sum_n b_n p\big(t - (2n-1)T_b\big) \quad (2.8\text{-}31)$$

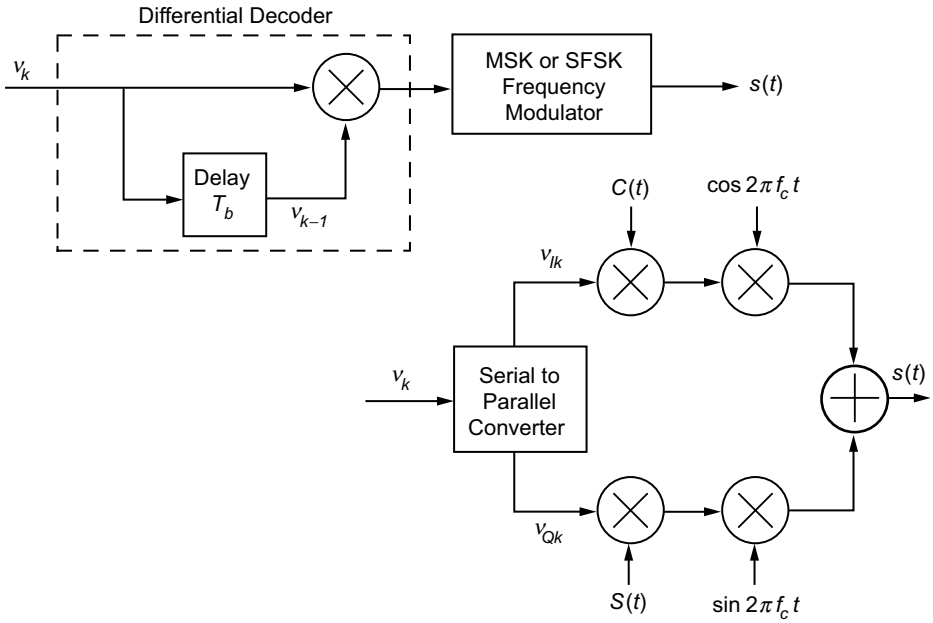

Fig. 2-14. CPM and equivalent I-Q implementations of precoded MSK or SFSK.

the PSD is given by [1, Chap. 2]

$$S_s(f) = \frac{1}{4}\left[G(f - f_c) + G(f + f_c)\right] \tag{2.8-32}$$

where $G(f)$ is the equivalent baseband PSD and is related to the PSD, $S_m(f)$, of $m_I(t)$ or $m_Q(t)$ by

$$G(f) = 2A^2 S_m(f), \quad S_m(f) = \frac{1}{2T_b}|P(f)|^2 \tag{2.8-33}$$

with $P(f)$ denoting the Fourier transform of the pulse shape $p(t)$. For MSK, we would have $A = \sqrt{2E_b/T_b}$ and $p(t)$ given by (2.8-25) with Fourier transform

$$P(f) = \frac{4T_b}{\pi} e^{-j2\pi f T_b} \frac{\cos 2\pi f T_b}{1 - 16f^2 T_b^2} \tag{2.8-34}$$

Substituting (2.8-34) in (2.8-33) gives the equivalent baseband PSD of MSK as

$$G(f) = \frac{32E_b}{\pi^2} \frac{\cos^2 2\pi f T_b}{\left(1 - 16 f^2 T_b^2\right)^2} \tag{2.8-35}$$

and the corresponding bandpass PSD as [1, Chap. 2]

$$S_s(f) = \frac{8E_b}{\pi^2} \left[\frac{\cos^2 2\pi (f - f_c) T_b}{\left(1 - 16 (f - f_c)^2 T_b^2\right)^2} + \frac{\cos^2 2\pi (f + f_c) T_b}{\left(1 - 16 (f + f_c)^2 T_b^2\right)^2} \right] \tag{2.8-36}$$

We observe from (2.8-35) that the main lobe of the lowpass PSD has its first null at $f = 3/4T_b$. Also, asymptotically for large f, the spectral sidelobes roll off at a rate f^{-4}. By comparison, the equivalent PSD of OQPSK wherein $A = \sqrt{E_b/T_b}$ and $p(t)$ is a unit amplitude rectangular pulse of duration $2T_b$, is given by

$$G(f) = 4E_b \frac{\sin^2 2\pi f T_b}{\left(2\pi f T_b\right)^2} \tag{2.8-37}$$

whose main lobe has its first null at $f = 1/2T_b$ and whose spectral sidelobes asymptotically roll off at a rate f^{-2}. Thus, we observe that while MSK (or precoded MSK) has a wider main lobe than OQPSK(or QPSK) by a factor of 3/2, its spectral sidelobes roll off at a rate two orders of magnitude faster. Figure 2-15 is an illustration of the normalized lowpass PSDs, $G(f)/2E_b$, of MSK and OQPSK obtained from (2.8-35) and (2.8-37), respectively, as well as that of SFSK, which is given by [1, Chap. 2]

$$G(f) =$$

$$2E_b \left[J_0\left(\frac{1}{4}\right) A_0(f) + 2\sum_{n=1}^{\infty} J_{2n}\left(\frac{1}{4}\right) B_{2n}(f) + 2\sum_{n=1}^{\infty} J_{2n-1}\left(\frac{1}{4}\right) B_{2n-1}(f) \right]^2,$$

$$A(f) = 2\frac{\sin 2\pi f T_b}{2\pi f T_b},$$

$$A_0(f) = \frac{1}{2} A\left(f + \frac{1}{4T_b}\right) + \frac{1}{2} A\left(f - \frac{1}{4T_b}\right) = \frac{4}{\pi} \frac{\cos 2\pi f T_b}{1 - 16 f^2 T_b^2},$$

$$A_{2n}(f) = \frac{1}{2}A\left(f + \frac{2n}{T_b}\right) + \frac{1}{2}A\left(f - \frac{2n}{T_b}\right),$$

$$A_{2n-1}(f) = \frac{1}{2}A\left(f + \frac{2n-1}{T_b}\right) - \frac{1}{2}A\left(f - \frac{2n-1}{T_b}\right),$$

$$B_{2n}(f) = \frac{1}{2}A_{2n}\left(f + \frac{1}{4T_b}\right) + \frac{1}{2}A_{2n}\left(f - \frac{1}{4T_b}\right),$$

$$B_{2n-1}(f) = -\frac{1}{2}A_{2n-1}\left(f + \frac{1}{4T_b}\right) + \frac{1}{2}A_{2n-1}\left(f - \frac{1}{4T_b}\right),$$

$$J_n(x) = n\text{th order Bessel function of the first kind} \tag{2.8-38}$$

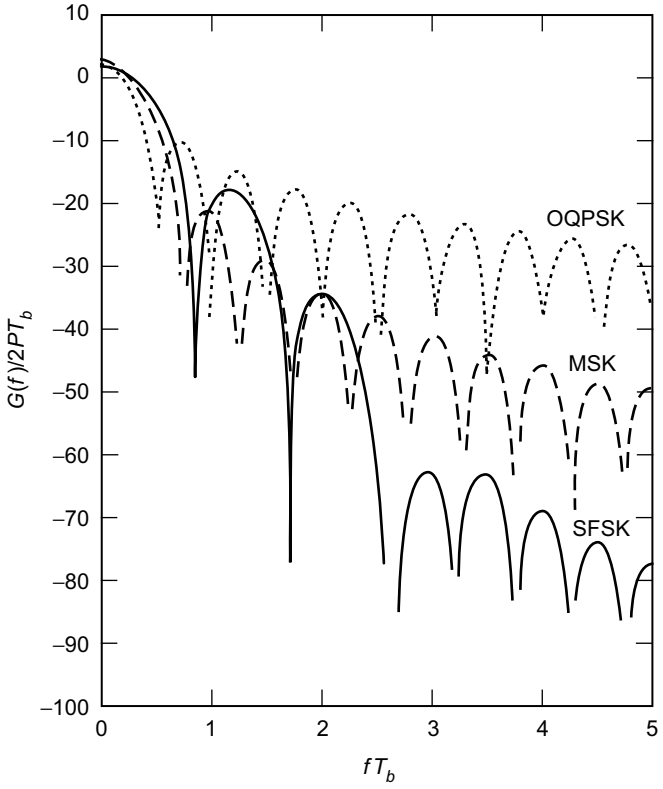

Fig. 2-15. A comparison of the equivalent baseband PSDs of MSK, OQPSK, and SFSK. Redrawn from [16].

whose main lobe is wider than that of MSK, but whose spectral sidelobes asymptotically roll off four orders of magnitude faster, i.e., at a rate f^{-8}. In fact, for the class of generalized MSK schemes, we can conclude that the smoother we make the shape of the frequency pulse, i.e., the more derivatives that go to zero at the endpoints $t = 0$ and $t = 2T_b$, the wider will be the main lobe but the faster the sidelobes will roll off.

Another way of interpreting the improved bandwidth efficiency that accompanies the equivalent I and Q pulse shaping is in terms of the fractional out-of-band power, defined as the fraction of the total power that lies outside a given bandwidth, i.e.,

$$\eta = 1 - \frac{\int_{-B/2}^{B/2} G(f)\, df}{\int_{-\infty}^{\infty} G(f)\, df} \tag{2.8-39}$$

Figure 2-16 is a plot of the fractional out-of-band power (in dB) versus BT_b for MSK, OQPSK, and SFSK, using the appropriate expression for $G(f)$ as determined from (2.8-35), (2.8-37), and (2.8-38), respectively.

2.8.1.5 Other Transmitter Representations.

a. Cross-Coupled I-Q Transmitter. A variation of the I-Q transmitter discussed in Sec. 2.8.1.2 is illustrated in Fig. 2-17 [24,25,26]. An modulated carrier at frequency f_c is multiplied by a lowpass sinusoidal signal at frequency $1/4T_b$ to produce a pair of unmodulated tones (carriers) at $f_2 = f_c + 1/4T_b$ and $f_1 = f_c - 1/4T_b$. These tones are separately extracted by narrow bandpass filters whose outputs, $s_1(t)$ and $s_2(t)$, are then summed and differenced to produce

$$\begin{aligned} z_c(t) = s_1(t) + s_2(t) &= \frac{1}{2}\cos\left[2\pi\left(f_c - \frac{1}{4T_b}\right)t\right] + \frac{1}{2}\cos\left[2\pi\left(f_c + \frac{1}{4T_b}\right)t\right] \\ &= \cos\left(\frac{\pi t}{2T_b}\right)\cos 2\pi f_c t \end{aligned} \tag{2.8-40}$$

$$\begin{aligned} z_s(t) = s_1(t) - s_2(t) &= \frac{1}{2}\cos\left[2\pi\left(f_c - \frac{1}{4T_b}\right)t\right] - \frac{1}{2}\cos\left[2\pi\left(f_c + \frac{1}{4T_b}\right)t\right] \\ &= \sin\left(\frac{\pi t}{2T_b}\right)\sin 2\pi f_c t \end{aligned}$$

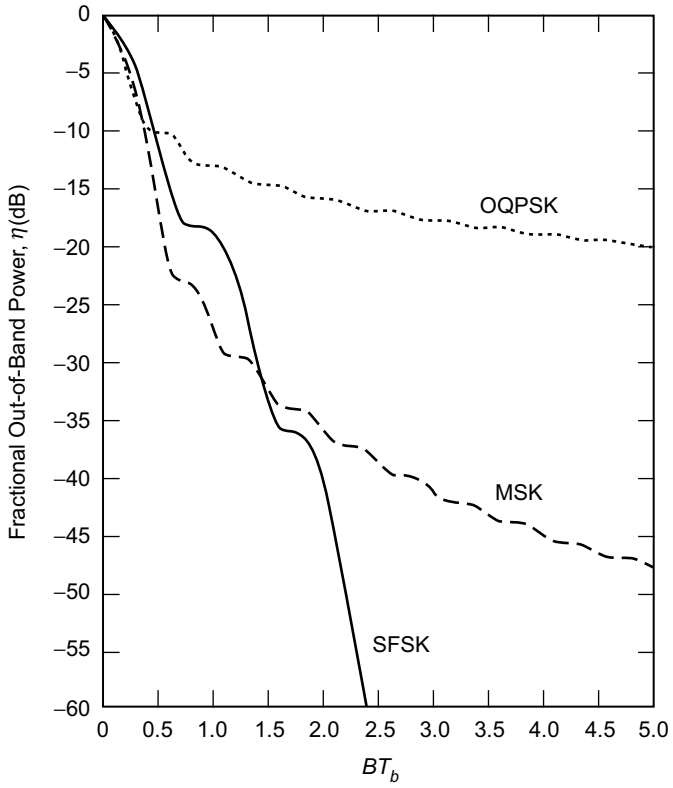

Fig. 2-16. A comparison of the fractional out-of-band power performance of MSK, OQPSK, and SFSK. Redrawn from [16].

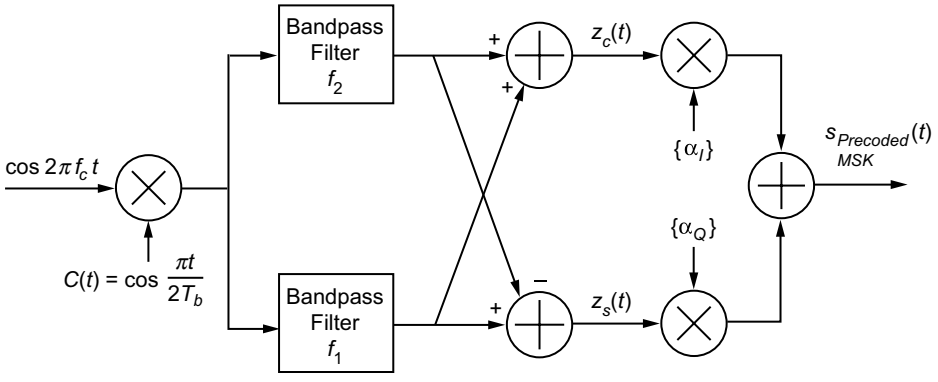

Fig. 2-17. Cross-coupled implementation of precoded MSK.

The signals $z_c(t)$ and $z_s(t)$ are respectively multiplied by I and Q data sequences $\{\alpha_I\}$ and $\{\alpha_Q\}$, each at a rate of $1/2T_b$ (and offset from each other by T_b s), and then differenced to produce the MSK (actually precoded MSK) output. The advantage of the implementation of Fig. 2-17 is that the signal coherence and the frequency deviation ratio are largely unaffected by variations in the data rate [25].

b. Rimoldi's Representation. As previously stated, the conventional CPM implementation of MSK produces a phase trellis that is symmetric about the horizontal axis, but that is time varying in that the possible phase states (reduced modulo 2π) alternate between $(0, \pi)$ and $(\pi/2, 3\pi/2)$ every T_b seconds. To remove this time-variation of the trellis, Rimoldi [27] demonstrated that CPM with a rational modulation index could be decomposed into the cascade of a memory encoder (finite-state machine) and a memoryless demodulator (signal waveform mapper). For the specific case of MSK, Rimoldi's transmitter is illustrated in Fig. 2-18. Imbalanced (0's and 1's) binary bits, $U_n = (1 - \alpha_n)/2$, are input to a memory one encoder. The current bit and the differentially encoded version of the previous bit (the encoder state) are used to define, via a binary-coded decimal (BCD) mapping, a pair of baseband signals (each chosen from a set of four possible waveforms) to be modulated onto I and Q carriers for transmission over the channel. Because of the imbalance of the data, the phase trellis is tilted as shown in Fig. 2-19, but on the other hand, it is now time invariant, i.e., the phase states (reduced modulo 2π) at all time instants (integer multiples of the bit time) are $(0, \pi)$. This transmitter implementation suggests the use of a simple two-state trellis decoder, which will be discussed in the next section

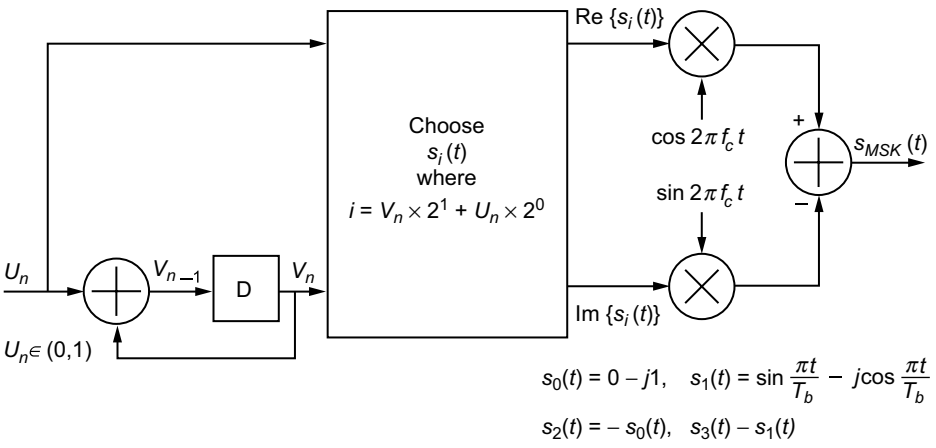

Fig. 2-18. MSK transmitter based on Rimoldi decomposition of CPM.

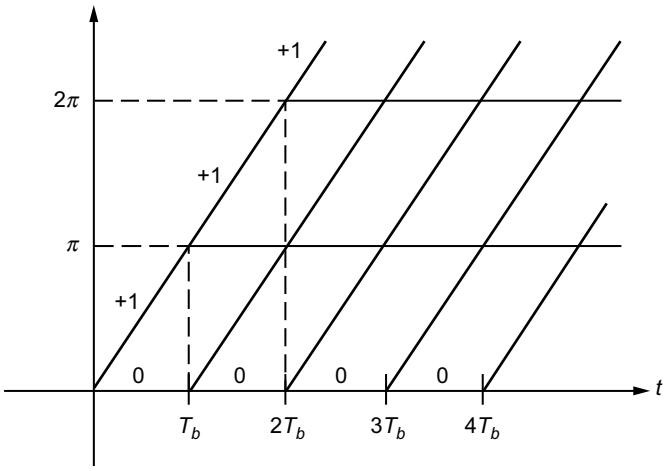

Fig. 2-19. Tilted (time-invariant) phase trellis for Rimoldi's MSK representation. Phase states (mod 2π) are $(0, \pi)$ for all n.

dealing with memory receiver structures. Also, later on in Chap. 4, we shall use Rimoldi's representation as the basis for developing bandwidth-efficient MSK-type modulations with memory greater than one under the constraint of finite decoding delay. Such modulations are not constrained to be constant envelope (rather, the transmitted signals are constrained to have equal energy) and thus, we defer our discussion of these schemes until that time.

Rimoldi's representation can also be used to implement precoded MSK. The appropriate transmitter is illustrated in Fig. 2-20.

2.8.1.6 Receiver Performance—Coherent Detection. Depending on the particular form used to represent the MSK signal, e.g., CPM, parallel I-Q, serial, etc., many different forms of receivers have been suggested in the literature for performing coherent detection. These various forms fall into two classes: structures based on a memoryless transmitter representation and structures based on a memory transmitter representation. As we shall see, all of these structures, however, are, themselves, memoryless.

a. Structures Based on a Memoryless Transmitter Representation. The most popular structure for coherent reception of MSK that is based on a memoryless transmitter representation corresponds to a parallel I-Q representation and has already been illustrated in Fig. 2-12. Here, the received signal plus noise is

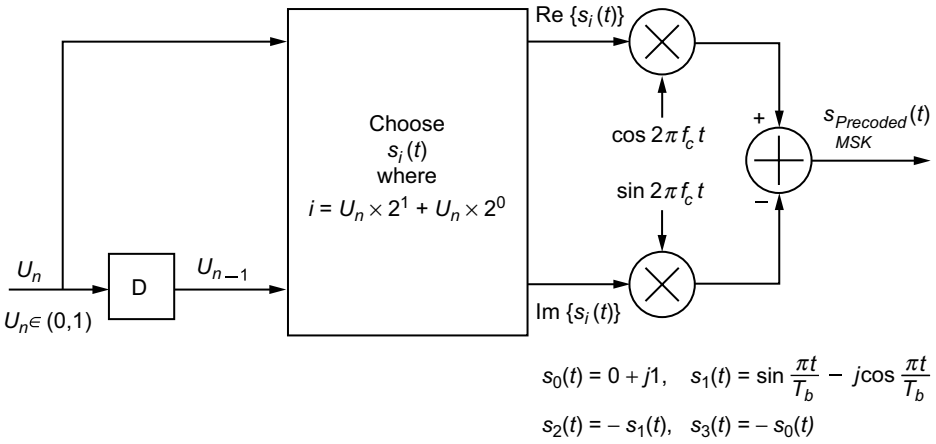

Fig. 2-20. Precoded MSK transmitter based on Rimoldi decomposition of CPM.

multiplied by the I and Q "carriers,"[11] $z_c(t)$ and $z_s(t)$, respectively, followed by integrate-and-dump (I&D) circuits of duration $2T_b$ seconds that are timed to match the zero crossings of the I and Q symbol waveforms. The multiplier-integrator combination constitutes a matched filter that, in the case of AWGN and no intersymbol interference (ISI), results in optimum detection. Means for producing the I and Q demodulation signals $z_c(t)$ and $z_s(t)$ will be discussed in the section on synchronization techniques.

b. Structures Based on a Memory Transmitter Representation. As noted in Sec. 2.8.1.5*b*, MSK (or precoded MSK) can be viewed as a cascade of a memory one encoder and a memoryless modulator. As such, a receiver can be implemented based on MLSE detection. For precoded MSK, the trellis diagram that appropriately represents the transitions between states is given in Fig. 2-21. Each branch of the trellis is labeled with the input bit (0 or 1) that causes a transition and the corresponding waveform (complex) that is transmitted as a result of that transition. The decision metrics based on a two-symbol observation that result in the surviving paths illustrated in Fig. 2-21 are

$$\int_{nT_b}^{(n+1)T_b} r(t)\, s_1(t)\, dt + \int_{(n+1)T_b}^{(n+2)T_b} r(t)\, s_0(t)\, dt > \int_{nT_b}^{(n+1)T_b} r(t)\, s_3(t)\, dt$$

$$+ \int_{(n+1)T_b}^{(n+2)T_b} r(t)\, s_1(t)\, dt \qquad (2.8\text{-}41a)$$

[11] The word "carrier" here is used to denote the combination (product) of the true carrier and the symbol waveform (clock).

and

$$\int_{nT_b}^{(n+1)T_b} r(t)\, s_1(t)\, dt + \int_{(n+1)T_b}^{(n+2)T_b} r(t)\, s_2(t)\, dt > \int_{nT_b}^{(n+1)T_b} r(t)\, s_3(t)\, dt$$

$$+ \int_{(n+1)T_b}^{(n+2)T_b} r(t)\, s_3(t)\, dt \tag{2.8-41b}$$

Noting from Fig. 2-20 that $s_3(t) = -s_0(t)$ and $s_2(t) = -s_1(t)$, (2.8-41a) and (2.8-41b) can be rewritten as

$$\int_{nT_b}^{(n+1)T_b} r(t)\, s_0(t)\, dt + \int_{(n+1)T_b}^{(n+2)T_b} r(t)\, s_0(t)\, dt > -\int_{nT_b}^{(n+1)T_b} r(t)\, s_1(t)\, dt$$

$$+ \int_{(n+1)T_b}^{(n+2)T_b} r(t)\, s_1(t)\, dt \tag{2.8-42a}$$

and

$$\int_{nT_b}^{(n+1)T_b} r(t)\, s_0(t)\, dt + \int_{(n+1)T_b}^{(n+2)T_b} r(t)\, s_0(t)\, dt > -\int_{nT_b}^{(n+1)T_b} r(t)\, s_1(t)\, dt$$

$$+ \int_{(n+1)T_b}^{(n+2)T_b} r(t)\, s_1(t)\, dt \tag{2.8-42b}$$

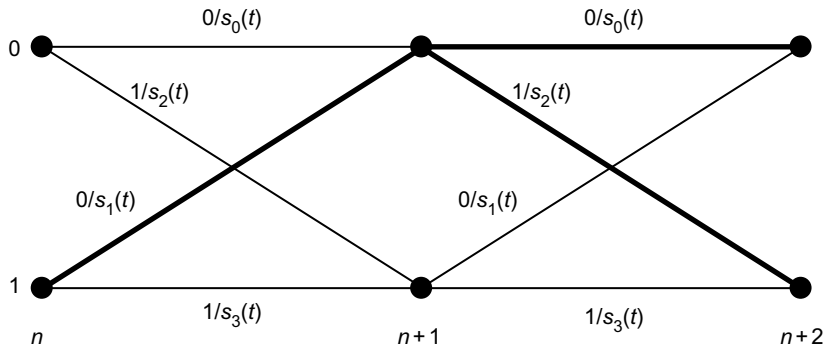

Fig. 2-21. A complex baseband trellis. Surviving paths for decoding $\hat{U}_n = 0$ in the interval $(n, n+1)$ assuming state "1" at time n are indicated by heavy lines.

which are identical and suggest the memoryless receiver illustrated in Fig. 2-22 [27].[12] Thus, we conclude that MSK (or precoded MSK) is a memory one type of trellis-coded modulation (TCM) that can be decoded with a finite (one bit) decoding delay, i.e., the decision on the nth bit can be made at the conclusion of observing the received signal for the n+1st transmission interval.

Massey [28] suggests an alternative representation of MSK (or precoded MSK) in the form of a single-input, two-output sequential transducer followed by an RF selector switch (Fig. 2-23). Analogous to the representation in (2.8-30), for precoded MSK, the sequential transducer implements the ternary sequences $\alpha_k^+ = (1/2)(\alpha_{k-1} + \alpha_k)$ and $\alpha_k^- = (-1)^k (1/2)(\alpha_{k-1} - \alpha_k)$. Note as before that α_k^+ is nonzero only when α_k^- is zero and vice versa. The function of the RF selector switch is to select one of the carriers for the signal to be transmitted in each bit interval according to the rule:

$$s(t) = \begin{cases} r_2(t) & \text{if } \alpha_k^+ = 1 \\ -r_2(t) & \text{if } \alpha_k^+ = -1 \\ r_1(t) & \text{if } \alpha_k^- = 1 \\ -r_1(t) & \text{if } \alpha_k^- = -1 \end{cases}, \quad r_i(t) = \sqrt{\frac{2E_b}{T_b}} \cos 2\pi f_i t, \quad i = 1, 2 \qquad (2.8\text{-}43)$$

which represents four mutually exclusive possibilities. This form of modulator has the practical advantage of not requiring addition of RF signals or RF filtering since there is no actual mixing of the carriers with the modulating signals.

Massey shows that, analogous to what is shown in Fig. 2-21, the output of the modulator can be represented by a trellis (Fig. 2-24), where again each branch is labeled with the input bit and the signal transmitted. Note that the trellis is time varying (the branch labels alternate with a period of two). In view of the trellis representation in Fig. 2-24, the optimum receiver is again an MLSE that has the identical structure as that in Fig. 2-22, where the complex demodulation signals $s_0(t - (n+1)T_b)$ and $s_1(t - (n+1)T_b)$ are replaced by the real carriers $r_1(t)\, r_1(t)$ and $r_2(t)$ of (2.8-43), the real part of the comparator (difference) output is omitted, and the decision device outputs balanced $+1, -1$ data rather than $0, 1$ data.

Regardless of the particular receiver implementation employed, the BEP performance of ideal coherent detection[13] of MSK is given by

---

[12] It can be shown that the surviving paths corresponding to being in state "0" at time n leads to the identical decision metric as that in (2.8-41a) or (2.8-41b).

[13] By "ideal coherent detection," we mean a scenario wherein the local supplied carrier reference is perfectly phase (and frequency) synchronous with the received signal carrier. Later on, we explore the practical implications of imperfect carrier synchronization.

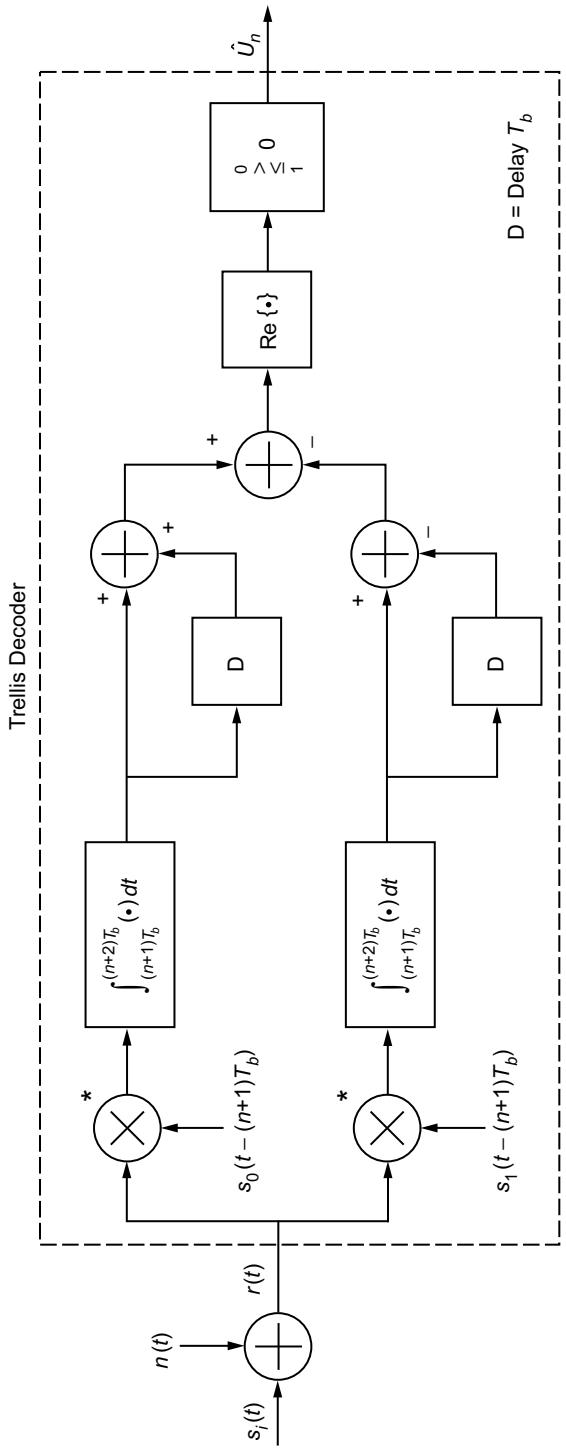

Fig. 2-22. Complex MLSE receiver.

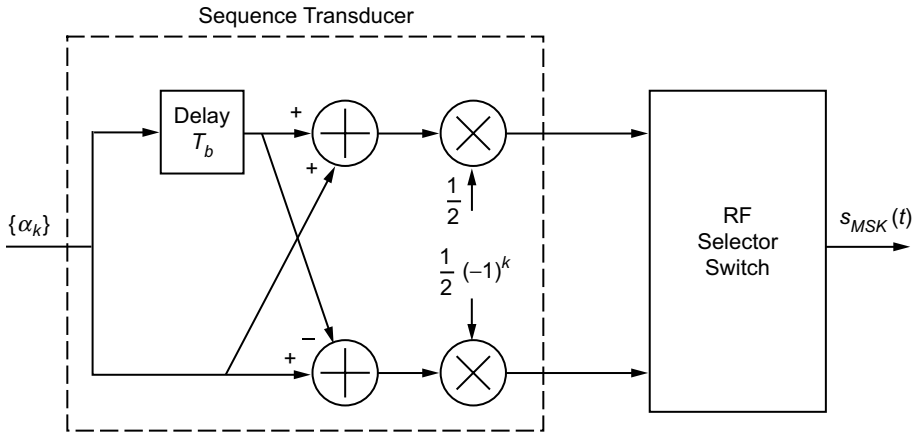

Fig. 2-23. Massey's precoded MSK transmitter.

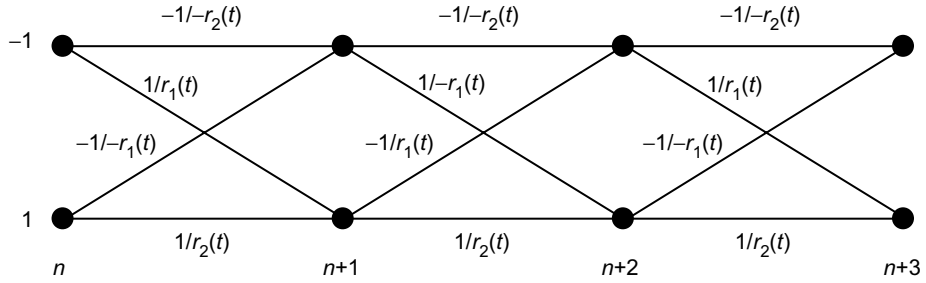

Fig. 2-24. Transmitter output trellis diagram.

$$P_b(E) = \text{erfc}\sqrt{\frac{E_b}{N_0}}\left(1 - \frac{1}{2}\text{erfc}\sqrt{\frac{E_b}{N_0}}\right) \tag{2.8-44}$$

whereas the equivalent performance of precoded MSK is

$$P_b(E) = \frac{1}{2}\text{erfc}\sqrt{\frac{E_b}{N_0}} \tag{2.8-45}$$

which is identical to that of ideal coherent detection of BPSK, QPSK, or OQPSK [see (2.6-2)]. Comparing (2.8-44) with (2.8-45), we observe that the former can be written in terms of the latter as

$$P_b(E)|_{\text{MSK}} = 2P_b(E)\Big|_{\substack{\text{precoded}\\ \text{MSK}}} \left(1 - P_b(E)\Big|_{\substack{\text{precoded}\\ \text{MSK}}}\right) \tag{2.8-46}$$

which reflects the penalty associated with the differential encoding/decoding operation inherent in MSK but not in precoded MSK as previously discussed. At a BEP of 10^{-5}, this amounts to a penalty of approximately a factor of two in error probability or equivalently a loss of 0.75 dB in E_b/N_0.

2.8.1.7 Receiver Performance—Differentially Coherent Detection. In addition to coherent detection, MSK can be differentially detected [29], as illustrated in Fig. 2-25. The MSK signal plus noise is multiplied by itself delayed one bit and phase shifted 90 deg. The resulting product is passed through a lowpass zonal filter that removes second harmonics of the carrier frequency terms. Also assumed is that the carrier frequency and data rate are integer related, i.e., $f_c T_b = k$, with k integer. Assuming that the MSK signal input to the receiver is in the form of (2.8-1) combined with (2.8-12), i.e.,

$$s(t) = \sqrt{\frac{2E_b}{T_b}}\cos\left(2\pi f_c t + \alpha_n \frac{\pi}{2T_b}t + x_n\right) = \sqrt{\frac{2E_b}{T_b}}\cos\Phi(t,\boldsymbol{\alpha}),$$

$$nT_b \le t \le (n+1)T_b \tag{2.8-47}$$

then the differential phase $\Delta\Phi \triangleq \Phi(t,\boldsymbol{\alpha}) - \Phi(t - T_b, \boldsymbol{\alpha})$ is given by

$$\Delta\Phi \stackrel{\Delta}{=} -(\alpha_{n-1} - \alpha_n)\frac{\pi}{2}\left(\frac{t}{T_b} - k\right) + \alpha_{n-1}\frac{\pi}{2} \tag{2.8-48}$$

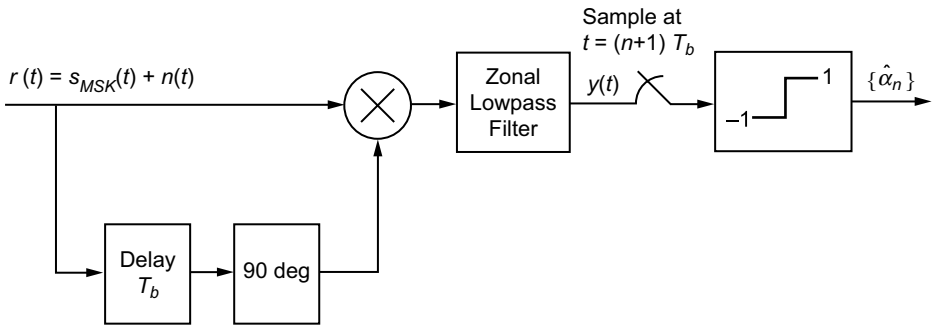

Fig. 2-25. Differentially coherent MSK receiver.

where we have made use of the phase continuity relation in (2.8-15) in arriving at (2.8-48). The mean of the lowpass zonal filter output can be shown to be given by

$$\overline{y(t)} = s(t)\, s_{90}(t) = \frac{E_b/T_b}{2} \sin \Delta\Phi \tag{2.8-49}$$

where the "90" subscript denotes a phase shift of 90 deg in the corresponding signal. Combining (2.8-48) and (2.8-49), the sampled mean of the lowpass zonal filter output at time $t = (n+1)\,T_b$ becomes

$$\overline{y((k+1)\,T_b)} = \frac{E_b/T_b}{2} \sin\left(\alpha_k \frac{\pi}{2}\right) = \alpha_k \frac{E_b/T_b}{2} \tag{2.8-50}$$

which clearly indicates the appropriateness of a hard limiter detector in the presence of noise. Figure 2-26 is an illustration of the various waveforms present in the differentially coherent receiver of Fig. 2-25 for a typical input data sequence.

2.8.1.8 Synchronization Techniques. In our discussion of coherent reception in Sec. 2.8.1.6, we implicitly assumed that a means was provided in the receiver for synchronizing the phase of the local demodulation reference(s) with that of the received signal carrier and also for time synchronizing the I&D circuits. Here we discuss several options for implementing such means.

One form of combined carrier and clock recovery that is synergistic with the transmitter form in Fig. 2-17 was originally proposed by DeBuda [30,31].[14] With reference to Fig. 2-27, the received MSK signal is first squared to produce an FSK signal at twice the carrier frequency and with twice the modulation index, i.e., $h = 1$, which is known as Sunde's FSK [32]. Whereas the MSK signal has no discrete (line) spectral components, after being squared, it has strong spectral components at $2f_1$ and $2f_2$, which can be used for synchronization. In fact, Sunde's FSK has 50 percent of its total power in these two line components (the other 50 percent of the total power is in a discrete line component at dc). To demonstrate this transformation from continuous to discrete spectrum, we square the MSK signal form in (2.8-30), which gives

[14] DeBuda also referred to MSK, in conjunction with his self-synchronizing circuit, as "fast FSK (FFSK)," which at the time was the more popular terminology in Canada.

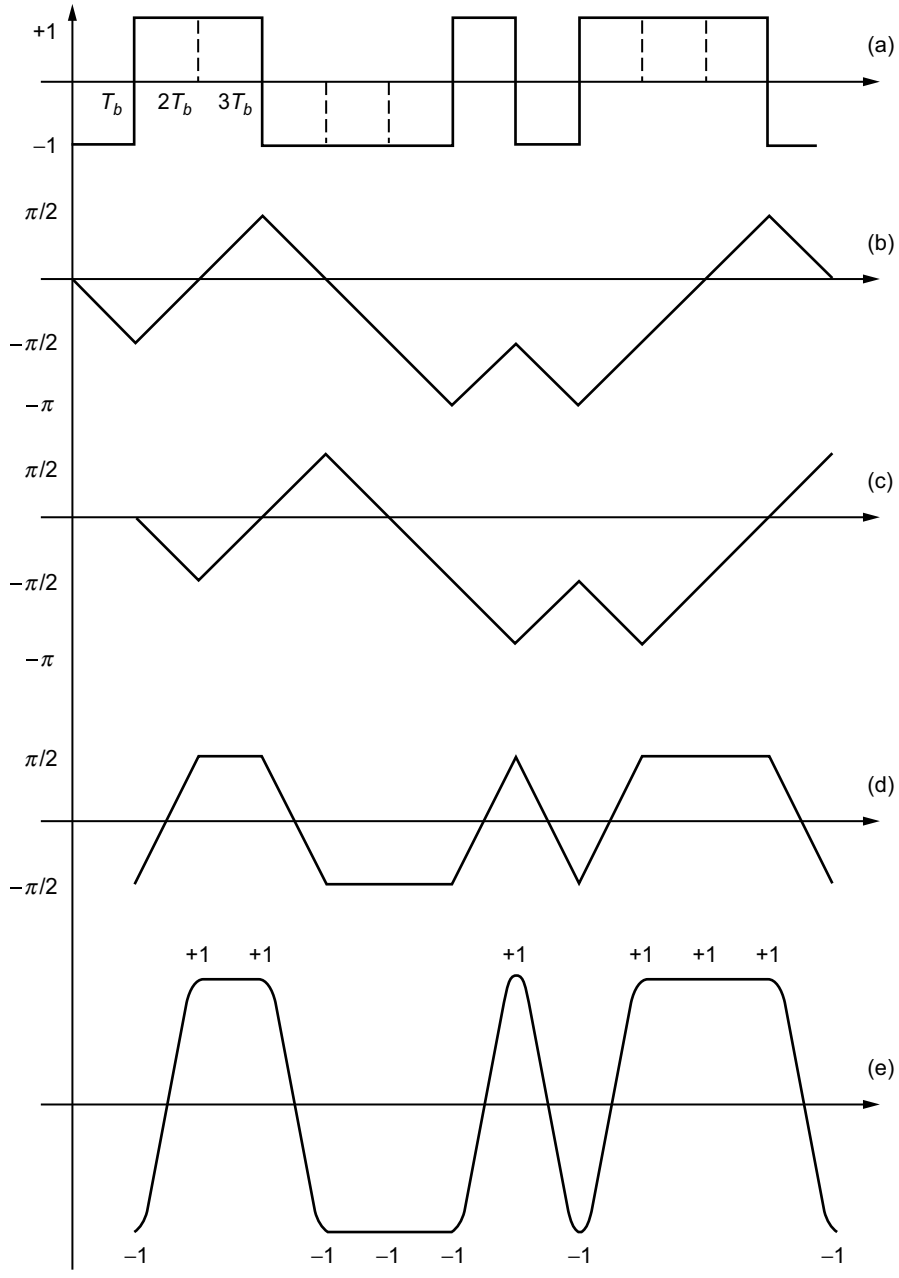

Fig. 2-26. Various waveforms present in the differentially coherent receiver shown in Fig. 2-25: (a) transmitted bit sequence, (b) transmitted phase, (c) transmitted phase delayed, (d) difference phase, and (e) multiplier output (sine of difference phase).

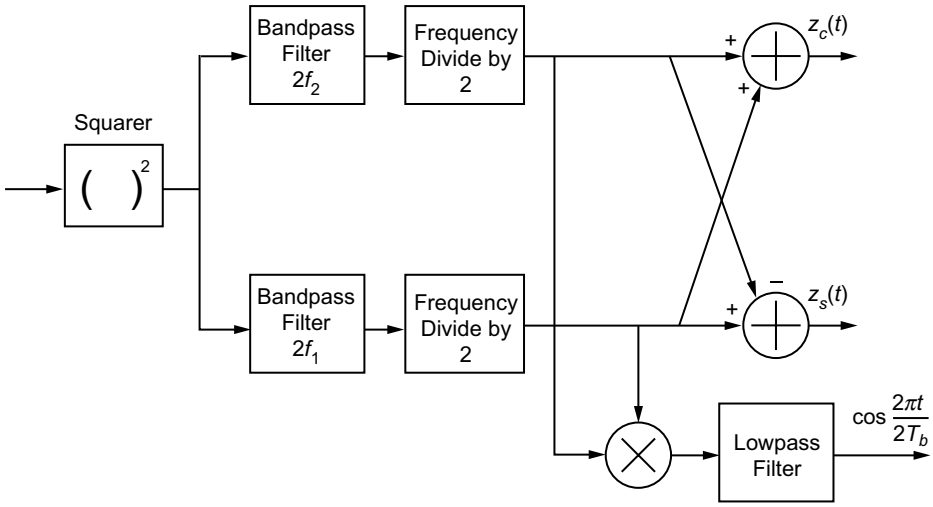

Fig. 2-27. DeBuda's carrier and symbol synchronization scheme.

$$s^2_{\mathrm{MSK}}(t) = \frac{2E_b}{T_b}\left[\left(v_n^+\right)^2\cos^2 2\pi f_2 t + \left(v_n^-\right)^2\cos^2 2\pi f_1 t + 2v_n^+ v_n^-\cos 2\pi f_2 t\cos 2\pi f_1\right]$$

$$= \frac{2E_b}{T_b}\left[\frac{1}{2} + \frac{1}{2}\left(v_n^+\right)^2\cos 4\pi f_2 t + \frac{1}{2}\left(v_n^-\right)^2\cos 4\pi f_1 t\right],$$

$$v_n^+ = \frac{v_{n-1}+v_n}{2}, \quad v_n^- = (-1)^n\left(\frac{v_{n-1}-v_n}{2}\right) \qquad (2.8\text{-}51)$$

where we have made use of the fact that since either v_n^+ or v_n^- is always equal to zero, then $v_n^+ v_n^- = 0$. Also, either $\left(v_n^+\right)^2 = 1$ and $\left(v_n^-\right)^2 = 0$ or vice versa, which establishes (2.8-51) as a signal with only discrete line components. The components at $2f_1$ and $2f_2$ are extracted by bandpass filters (in practice, phase-locked loops) and then frequency divided to produce $s_1(t) = (1/2)\cos 2\pi f_1 t$ and $s_2(t) = (1/2)\cos 2\pi f_2 t$. The sum and difference of these two signals produce the reference "carriers" $z_c(t) = C(t)\cos 2\pi f_c t$ and $z_s(t) = S(t)\sin 2\pi f_c t$, respectively, needed in Fig. 2-12. Finally, multiplying $s_1(t)$ and $s_2(t)$ and lowpass filtering the result produces $(1/8)\cos 2\pi t/2T_b$ (a signal at 1/2 the bit rate), which provides the desired timing information for the I&Ds in Fig. 2-12.

Another joint carrier and timing synchronization scheme for MSK was derived by Booth [33] in the form of a closed loop motivated by the maximum a posteriori (MAP) estimation of carrier phase and symbol timing. The resulting

structure [Fig. 2-28(a)] is an overlay of two MAP estimation I-Q closed loops—one typical of a carrier synchronization loop, assuming known symbol timing [Fig. 2-28(b)] and one typical of a symbol timing loop, assuming known carrier phase [Fig. 2-28(c)]. In fact, the carrier synchronization component loop is identical to what would be obtained for sinusoidally pulse-shaped OQPSK.

Finally, many other synchronization structures have been developed for MSK and conventional (single modulation index) binary CPM, which, by definition, would also be suited to MSK. A sampling of these is given in Refs. 34–40. In the interest of brevity, however, we do not discuss these here. Instead, the interested reader is referred to the cited references for the details.

2.8.2 Partial Response—Gaussian MSK

GMSK was first introduced by Murota, Kinoshita, and Hirada [41] in 1981 as a highly bandwidth-efficient constant envelope modulation scheme for communication in the 900-MHz land mobile radio environment (see [42,43] for field experimental results of performance in this frequency band). In simple terms, GMSK is an $h = 0.5$ partial-response CPM scheme obtained by filtering the rectangular frequency pulses characteristic of MSK with a filter having a Gaussian impulse response prior to frequency modulation of the carrier.[15] As such, the GMSK frequency pulse is the difference of two time-displaced (by T_b seconds) Gaussian probability integrals (Q-functions), i.e.,[16]

$$g(t) = \frac{1}{2T_b}\left[Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}}\left(\frac{t}{T_b} - 1\right)\right) - Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}}\frac{t}{T_b}\right)\right],$$

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy, \quad -\infty \le t \le \infty \qquad (2.8\text{-}52)$$

[15] It is important to emphasize that although the acronym GMSK was assigned to the term *Gaussian-filtered MSK* in [41], the modulation actually described in this reference applies the Gaussian filtering at baseband, i.e., prior to modulation onto the carrier, and, hence, it does not destroy the constant envelope property of the resulting modulation. Perhaps because of this poor usage of the term *Gaussian-filtered MSK*, occasionally there appears in the literature [44, p. 519] a misleading statement alluding to the fact that GMSK is an "MSK modulated signal passed through a Gaussian filter . . .," which would imply Gaussian filtering at RF, thereby destroying the constant envelope nature of the signal. This interpretation is not in keeping with the original description of GMSK in [41] and the large number of references that followed; thus, we caution the reader against adopting this usage.

[16] We assume here a frequency pulse shape, $g(t)$, that results from excitation of the Gaussian filter (arbitrarily assumed to have zero group delay) with the unit rectangular pulse $p(t) = 1, 0 \le t \le T_b$.

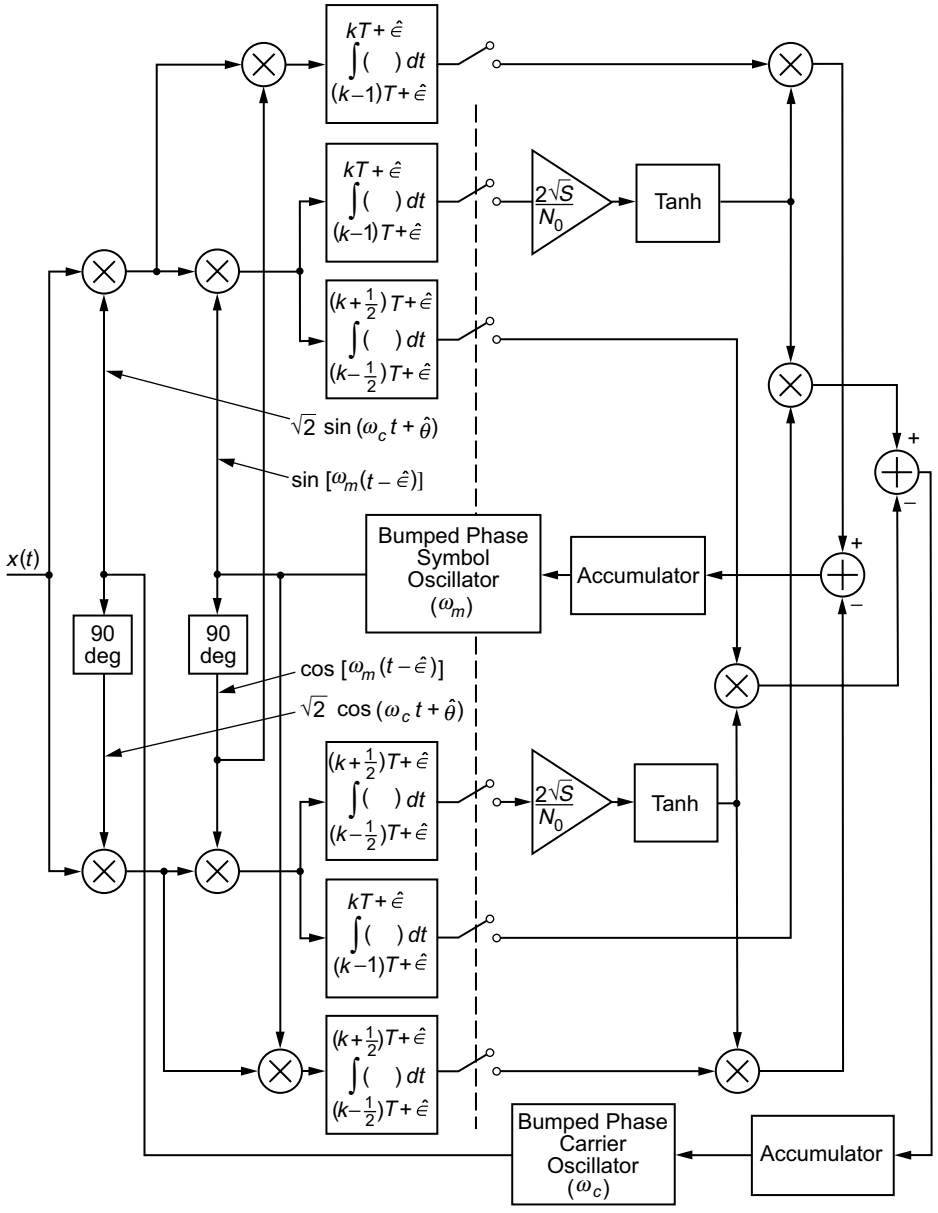

Fig. 2-28(a). Joint carrier and symbol MAP estimation loop for MSK modulation.

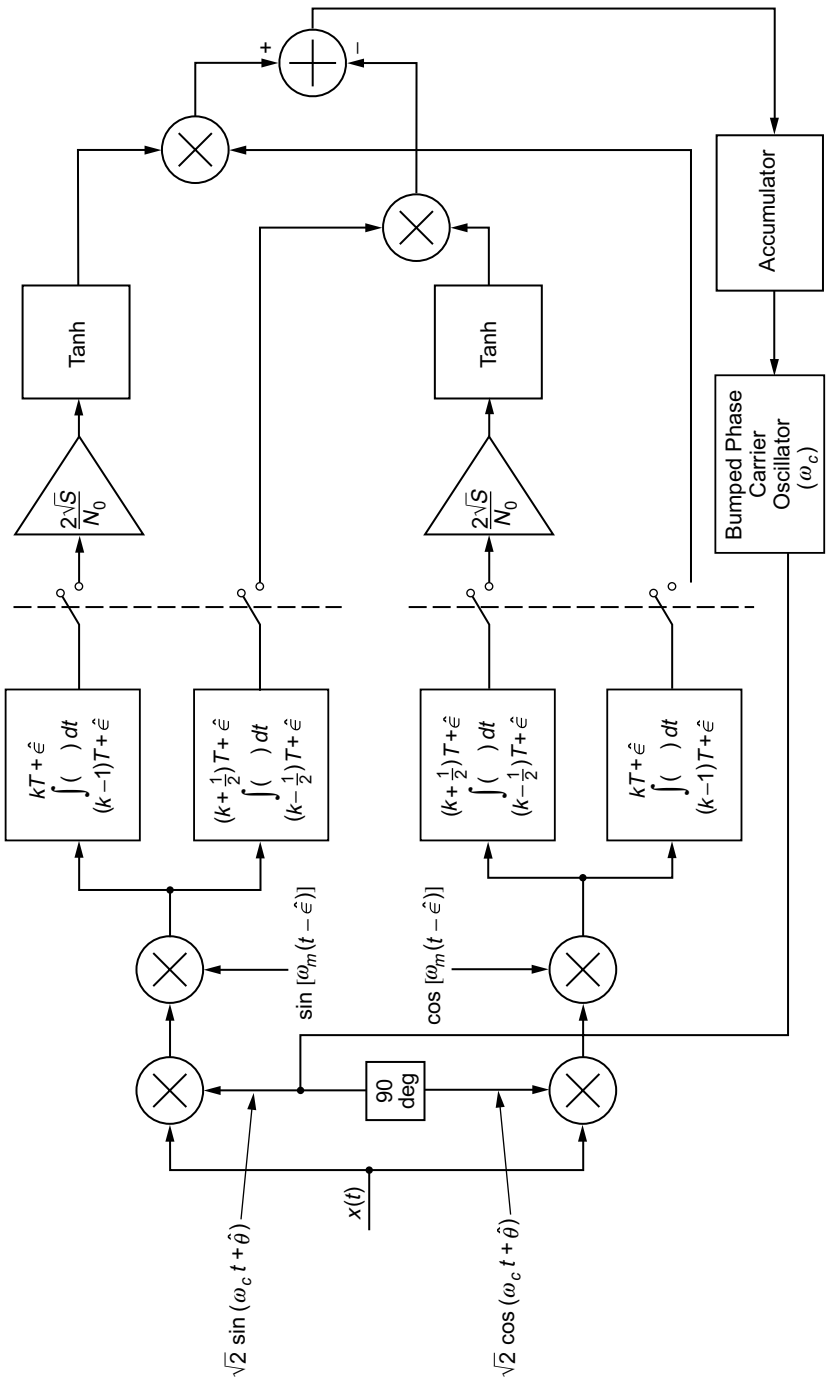

Fig. 2-28(b). Joint carrier and symbol MAP estimation loop for MSK modulation (carrier synchronization component).

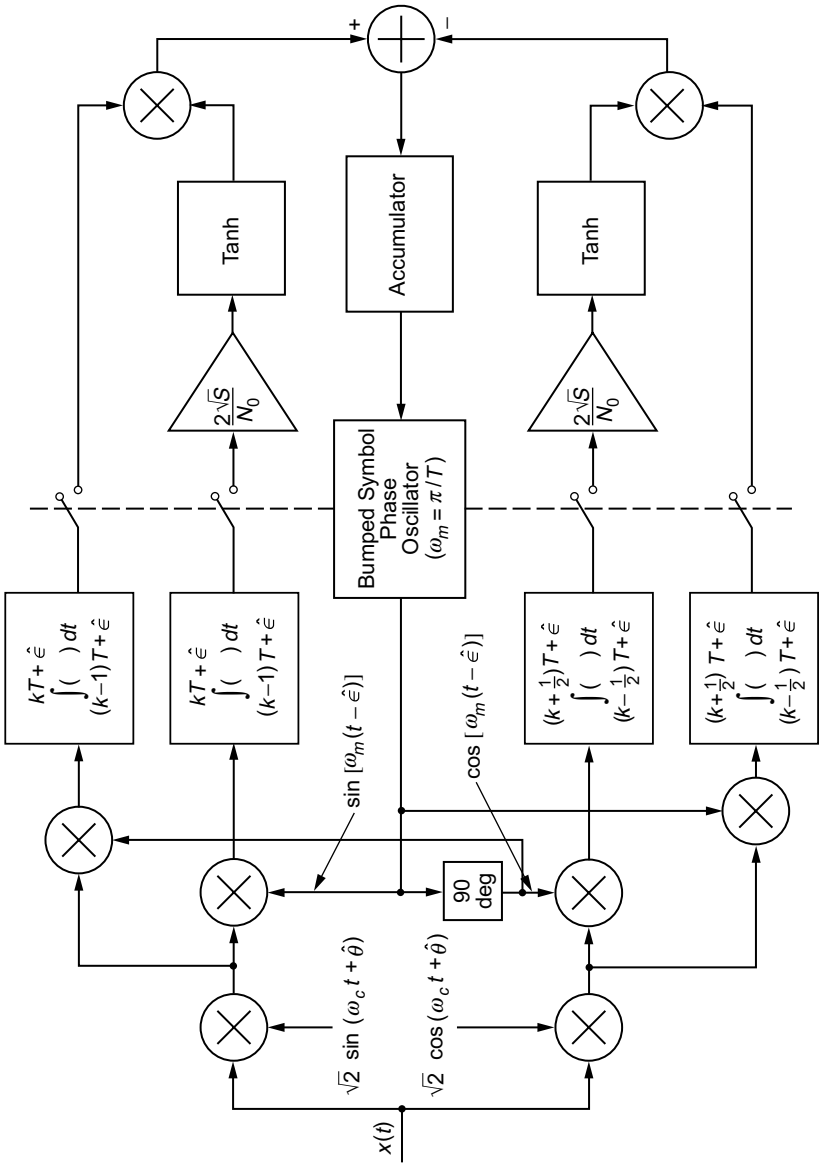

Fig. 2-28(c). Joint carrier and symbol MAP estimation loop for MSK modulation (symbol synchronization component).

where B is the 3-dB bandwidth of the lowpass Gaussian filter and is related to the noise bandwidth, B_N, of this filter by [45, Eq. (14)]

$$\frac{B}{B_N} = 2\sqrt{\frac{\ln 2}{\pi}} = 0.93944 \tag{2.8-53}$$

Smaller values of BT_b lead to a more compact spectrum but also introduce more ISI and, therefore, a degraded error probability performance. Thus, for a given application, the value of BT_b is selected as a compromise between spectral efficiency and BEP performance.

Since the Gaussian Q-function is doubly infinite in extent, it is common practice to time-truncate the GMSK frequency pulse so as to deal with finite ISI. For $BT_b = 0.25$, truncating $g(t)$ of (2.8-52) to four bit intervals is appropriate [46] whereas for $BT_b = 0.3$, the value used in the Global System for Mobile (GSM) application [47], considering ISI only from adjacent bits (i.e., time truncation to three bit intervals) has been shown to be sufficient [48]. Thus, in practical GMSK implementations, one employs the approximation (see Fig. 2-29)

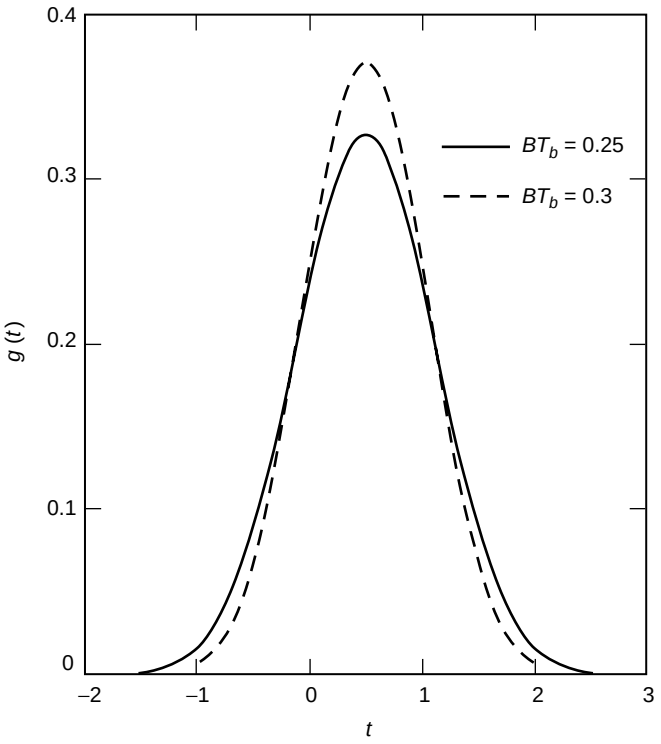

Fig. 2-29. GMSK frequency pulse.

$$g(t) = \begin{cases} \dfrac{1}{2T_b}\left[Q\left(\dfrac{2\pi BT_b}{\sqrt{\ln 2}}\left(\dfrac{t}{T_b}-1\right)\right) - Q\left(\dfrac{2\pi BT_b}{\sqrt{\ln 2}}\dfrac{t}{T_b}\right)\right], & -(L-1)\,T_b/2 \le t \\ & \le (L+1)\,T_b/2 \\ 0, & \text{otherwise} \end{cases} \tag{2.8-54}$$

where L is chosen as above in accordance with the value of BT_b.[17] Also, although $g(t)$ of (2.8-54) appears to have a "Gaussian-looking" shape, we emphasize that the word *Gaussian* in GMSK refers to the *impulse* response of the filter through which the input rectangular pulse train is passed and not the shape of the resulting frequency pulse.

2.8.2.1 Continuous Phase Modulation Representation. Based on the above, the CPM representation of GMSK is, analogous to (2.8-10),

$$s_{\text{MSK}}(t) = \sqrt{\frac{2E_b}{T_b}}\cos\left(2\pi f_c t + \frac{\pi}{2T_b}\sum_i \alpha_i \int \left[Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}}\left(\frac{\tau}{T_b}-(i+1)\right)\right)\right.\right.$$

$$\left.\left. -Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}}\left(\frac{\tau}{T_b}-i\right)\right)\right] d\tau\right), \quad nT_b \le t \le (n+1)\,T_b \tag{2.8-55}$$

which is implemented, analogous to Fig. 2-7, in Fig. 2-30(a). Equivalently, if the input is represented by its equivalent NRZ data stream (i.e., the frequency pulse stream that would ordinarily be inputted to the FM modulator in MSK), then the filter impulse response, $h(t)$, becomes Gaussian, as implied by the GMSK acronym, i.e.,

$$h(t) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{t^2}{2\sigma^2}\right), \quad \sigma^2 = \frac{\ln 2}{(2\pi B)^2} \tag{2.8-56}$$

(appropriately time-truncated as discussed above), and the implementation appears as in Fig. 2-30(b).

---

[17] Technically speaking, $g(t)$ of (2.8-53) should be scaled by a constant C so as to satisfy a condition analogous to (2.8-5), namely,

$$q(t) = \int_{-\infty}^{t} g(\tau)d\tau = \begin{cases} 0, & t \le -(L-1)\,T_b/2 \\ 1/2, & t \ge (L+1)\,T_b/2 \end{cases}$$

However, for the values of BT_b of practical interest, i.e., $BT_b \ge 0.25$, the scaling constant is ignored, i.e., C is nominally taken as unity.

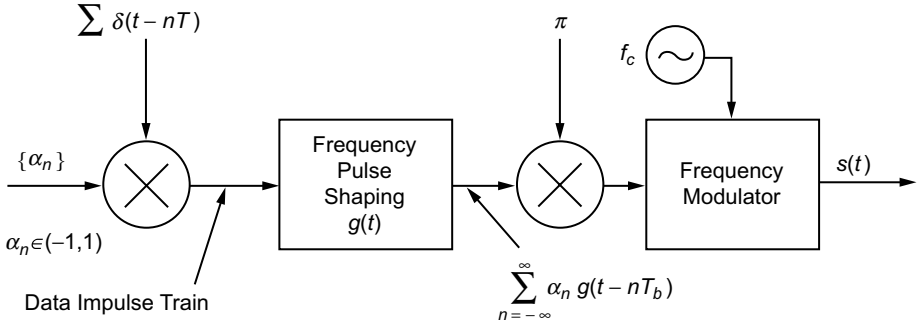

Fig. 2-30(a). GMSK transmitter (CPM representation).

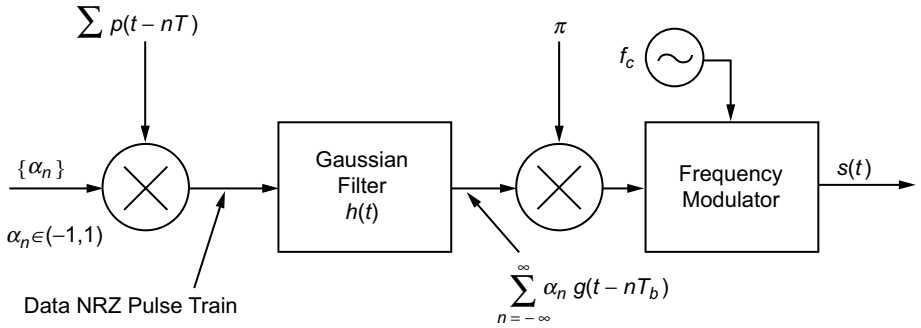

Fig. 2-30(b). Equivalent GMSK transmitter (CPM representation).

The frequency modulator in Fig. 2-30(a) or 2-30(b) is typically implemented with a phase-locked loop (PLL) synthesizer whose voltage-controlled oscillator (VCO) input is the point at which the modulation is injected. When long strings of zeros or ones are present in the data, the spectrum of the modulation extends to dc, which presents a problem, since PLL frequency synthesizers implemented as above do not respond to this low-frequency signal due to their inherent high-pass filter characteristic. As such, the VCO output (the location of the modulated signal) would not contain the low-frequency content of the information (modulating) signal. By contrast, if the modulation were to be injected at the input of the master oscillator preceding the PLL (the oscillator must be capable of being modulated by a voltage signal), then since this oscillator is not in the loop, the VCO output would contain the low-frequency content of the modulation (i.e., that within the loop filter bandwidth) but not its high-frequency content. Clearly then, a combination of these two approaches would yield the desirable result of constant modulation sensitivity, irrespective of the loop bandwidth. Such an FM scheme is referred to as two-point modulation [49] and corresponds to a

dc-coupled GMSK modulator wherein the Gaussian filtered input signal is split sending one portion to the VCO modulation input and the other to the PLL master oscillator input.

2.8.2.2 Equivalent I-Q Representations. For high carrier frequencies, direct synthesis of the GMSK signal as in Fig. 2-7, using a digital approach is impractical since maintaining an adequate sampling rate requires an extremely high operating frequency. Instead, one can resort to a quadrature implementation where lowpass I and Q signals containing the phase information are generated that vary much slower than the phase of the modulated carrier, thus making it feasible to implement them digitally. Applying the simple trigonometric rule for the cosine of the sum of two angles to (2.8-55), we obtain

$$s_{\mathrm{MSK}}(t) = \sqrt{\frac{2E_b}{T_b}}\left[\cos\phi(t,\boldsymbol{\alpha})\cos 2\pi f_c t - \sin\phi(t,\boldsymbol{\alpha})\sin 2\pi f_c t\right] \tag{2.8-57}$$

where

$$\phi(t,\boldsymbol{\alpha}) = \frac{\pi}{2T_b}\sum_i \alpha_i \int \left\{ Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}}\left(\frac{\tau}{T_b} - (i+1)\right)\right) - Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}}\left(\frac{\tau}{T_b} - i\right)\right)\right\} d\tau \tag{2.8-58}$$

Conceptually then, an I-Q receiver for GMSK is one that performs the following sequence of steps: first, the Gaussian-filtered NRZ data stream is generated. Next, integration is performed to produce the instantaneous phase of (2.8-58). Finally, the integrator output is passed through sine and cosine read-only memories (ROMs) whose outputs are applied to I and Q carriers (see Fig. 2-31). Such a scheme has also been referred to as quadrature cross-correlated GMSK (see [50, Fig. 4.3.20] for an illustration similar to Fig. 2-31). Several commercial vendors and industrial organizations, e.g., Alcatel and Aerospace, have digitally implemented this generic approach in the transmitter design of their GMSK modems. In these implementations, the block labeled "Gaussian filter" is either an actual filter that approximates the Gaussian impulse response as per (2.8-54) or, more efficiently, a ROM table lookup, whereas the block labeled "integrator" is typically performed by a "phase accumulator."[18]

[18] Without loss in generality, the Gaussian filter and integrator blocks can be switched as is the case in some of the implementations.

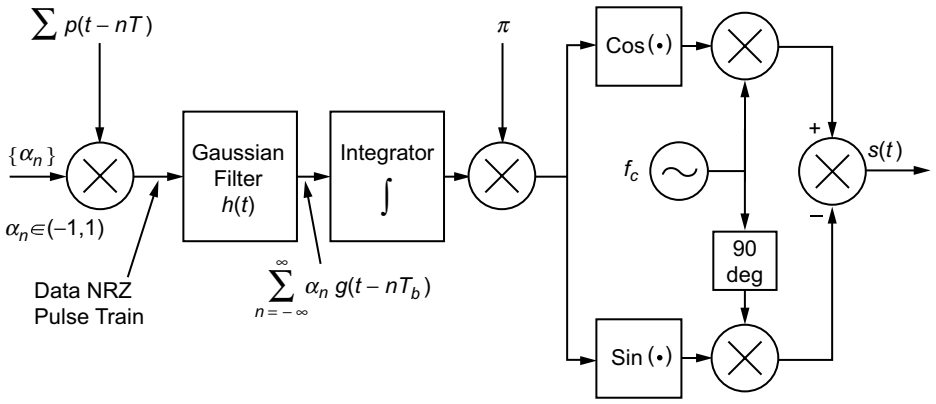

Fig. 2-31. GMSK transmitter (I-Q representation).

In [45], an efficient I-Q implementation of a GMSK modulator is presented that skips the above sequence of steps and instead generates the I and Q baseband signals directly from the binary data, thereby eliminating the errors in filtering, phase truncation, and sine/cosine computation inherent in the conventional architecture. A brief description of this method is as follows, based on the assumption of ISI only from adjacent symbols, i.e., $L = 3$.

Consider the GMSK frequency response (pulse train) that generates the phase of (2.8-58). If we impose the condition that this response in the mth bit interval, $mT_b \le |t| \le (m+1)\,T_b$, be dependent only on the bit of interest, α_m, and its two nearest neighbors, α_{m-1} and α_{m+1}, i.e., only adjacent ISI, then it can be shown [45, Eqs. (28), (29)] that it is sufficient to require

$$\left.\begin{aligned} Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}}\right) &\cong 0 \\ Q\left(-\frac{2\pi BT_b}{\sqrt{\ln 2}}\right) &\cong 1 \end{aligned}\right\} \tag{2.8-59}$$

Assuming (2.8-59) is true, then since by superposition the response to a train of NRZ pulses varying from -1 to 1 is the equivalent to the response to a rectangular pulse train varying from 0 to 2 minus a constant of value 1, the normalized frequency response in the above interval can be expressed as

$$g_m(t) \triangleq \sum_{i=m-1,m,m+1} (\alpha_i + 1) \left[Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}} \left(\frac{t}{T_b} - (i+1) \right) \right) \right.$$

$$\left. - Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}} \left(\frac{t}{T_b} - i \right) \right) \right] dt - 1$$

$$\cong (\alpha_{m-1} + 1)\, Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}} \left(\frac{t}{T_b} - m \right) \right)$$

$$+ (\alpha_m + 1) \left[Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}} \left(\frac{t}{T_b} - (m+1) \right) \right) - Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}} \left(\frac{t}{T_b} - m \right) \right) \right]$$

$$+ (\alpha_{m+1} + 1) \left[1 - Q\left(\frac{2\pi BT_b}{\sqrt{\ln 2}} \left(\frac{t}{T_b} - (m+1) \right) \right) \right] - 1 \qquad (2.8\text{-}60)$$

Alternatively, since the Gaussian Q-function can be expressed in terms of the error function by $Q(x) = (1/2)\left[1 + \operatorname{erf}\left(x/\sqrt{2}\right)\right]$, then letting $\alpha_i' = (1/2)(\alpha_i + 1)$ denote the (0,1) equivalent of the $(-1, 1)$ α_i's, and introducing the constant $\beta \triangleq \pi B\sqrt{2/\ln 2}$, as in Eq. (19) of Ref. 45, (2.8-60) can be rewritten as

$$g_m(t) \cong \alpha_{m-1}' \left[1 - \operatorname{erf}\left(\beta(t - mT_b)\right)\right]$$

$$+ \alpha_m' \left[\operatorname{erf}\left(\beta(t - mT_b)\right) - \operatorname{erf}\left(\beta(t - (m+1)T_b)\right)\right]$$

$$+ \alpha_{m+1}' \left[1 + \operatorname{erf}\left(\beta(t - (m+1)T_b)\right)\right] - 1 \qquad (2.8\text{-}61)$$

Corresponding to the values (0,1) for each of the three α_i's in (2.8-61), there are eight possible waveforms $f_i(t - mT_b), i = 0, 1, 2, \cdots, 7$ that characterize the frequency response in the mth bit interval. These are given in Table 2-1 assuming $m = 0$ for simplicity.

Table 2-1. Possible frequency responses in the interval $0 \leq t \leq T_b$.

$\alpha'_{-1}, \alpha'_0, \alpha'_1$	i	$f_i(t)$
000	0	-1
001	1	$\operatorname{erf}\left(\beta(t-T_b)\right)$
010	2	$\operatorname{erf}(\beta t) - \operatorname{erf}\left(\beta(t-T_b)\right) - 1$
011	3	$\operatorname{erf}(\beta t)$
100	4	$-\operatorname{erf}(\beta t)$
101	5	$-\operatorname{erf}(\beta t) + \operatorname{erf}\left(\beta(t-T_b)\right) + 1$
110	6	$-\operatorname{erf}\left(\beta(t-T_b)\right)$
111	7	1

We observe from this table that there are only three independent frequency response waveforms, i.e., $f_2(t), f_3(t), f_7(t)$, in that the remaining five can be obtained from these three by means of simple operations, namely,

$$\left.\begin{aligned} f_0(t) &= -f_7(t) \\ f_1(t) &= f_3(t) - f_2(t) - f_7(t) = f_3(t - T_b) \\ f_4(t) &= -f_3(t) \\ f_5(t) &= -f_2(t) \\ f_6(t) &= -f_1(t) \end{aligned}\right\} \tag{2.8-62}$$

In view of the above, the frequency modulating signal corresponding to the phase modulating signal of (2.8-58) can be expressed in the form of a data-dependent pulse train as

$$f(t, \boldsymbol{\alpha}) = \frac{1}{2\pi}\frac{d}{dt}\phi(t, \boldsymbol{\alpha}) = \frac{1}{4T_b}\sum_i f_{l(i)}(t - iT_b)\, p(t - iT_b) \tag{2.8-63}$$

where as before, $p(t)$ is a unit amplitude rectangular pulse in the interval $0 \leq t \leq T_b$ and the index $l(i) = 4a_{i-1} + 2a_i + a_{i+1}$ is the decimal equivalent of

the 3-bit binary sequence influencing the ith bit interval and determines the particular frequency waveform for that interval in accordance with Table 2-1. The corresponding complex phase modulating signal can be written in the form

$$\exp\{\phi(t,\boldsymbol{\alpha})\} = \exp\left\{\sum_i \phi_{l(i)}(t - iT_b)\, p(t - iT_b)\right\},$$

$$\phi_i(t) = \frac{\pi}{2T_b}\int_0^t f_i(\tau)\, d\tau + \phi_i(0) \tag{2.8-64}$$

where $\phi_i(0)$ is the initial phase value that depends on the past history of the data sequence. Analogous to Table 2-1, there are eight possible phase responses in any given bit interval. These are evaluated in Ref. 45, using the approximation of (2.8-59) (reformulated in terms of the error function as $\text{erf}(\beta T_b) \cong 1$, $\text{erf}(-\beta T_b) \cong -1$), along with appropriate asymptotic expansions of the error function. Once again, there are only three independent phase response waveforms, e.g., $\phi_2(t), \phi_3(t), \phi_7(t)$, and the remaining five can be obtained from these three by means of simple operations.

The phase responses are used to determine phase trellises, keeping in mind that the sequences of possible phase trajectories generated by the 3-bit data sequences in each bit interval are constrained by the fact that only one of the 3 bits changes from interval to interval. Thus, for example, 010 can only be followed by 100 or 101. Furthermore, since we are interested only in the sine and cosine of the phase, it is sufficient to consider the phase trajectories modulo 2π. Using these trellises, it is shown in Ref. 45 that only four curves of T_b-s duration are needed to generate the I (from the cosine of the phase) and Q (from the sine of the phase) signals directly from the input data sequence for that bit interval. This is accomplished with a table lookup ROM that stores these four basic curves.

Finally, the practical trade-offs in terms of recent digital integrated circuit (IC) technology between the FM/VCO and I-Q transceiver architectures are discussed in Ref. 51.

2.8.2.3 Other GMSK Representations—The Laurent Expansion. A decade and a half ago, Laurent [52] described a representation for CPM in the form of a superposition of phase-shifted amplitude-modulation pulse (AMP) streams, the number of such being dependent on the amount of partial response in the modulation, as described by the duration (in bits) of the frequency pulse. A full-response scheme such as MSK required only a single pulse stream (with complex symbols). The primary focus of this work was on binary modulation

because of its relative simplicity of implementation.[19] The motivation for presenting such a representation was twofold. First, it allowed for easier evaluation of the autocorrelation and PSD of such modulations; in particular, simple results were specifically obtained for half-integer index modulations, i.e., ones whose frequency modulation index was of the form $h = n + 1/2$, n integer. Second, it allowed for approximation (with good accuracy) of CPM by a single pulse stream with one optimized pulse shape (called the "main pulse") and as such offered a synthesis means no more complicated than MSK.

Three years later, Kaleh [46] exploited Laurent's representation of CPM to allow for simple implementation of coherent receivers of such modulations, in particular, for the case of GMSK. Two forms of such receivers were considered, namely, a simplification of the optimum MLSE receiver and a linear MSK-type receiver, both of which yielded small degradation relative to the true optimum MLSE receiver.

In this section, we summarize the key results of these papers in so far as the transmitter implementation is concerned, devoting more time to the interpretations of the results than to the details of the derivations.

a. Exact AMP Representation of GMSK. In what follows, it will be convenient to deal with the complex envelope of the signal $s(t)$, i.e., the complex baseband signal, $\tilde{S}(t)$, defined by the relation

$$s(t) = \mathrm{Re}\left\{\tilde{S}(t)e^{j2\pi f_c t}\right\} \tag{2.8-65}$$

Thus, from (2.8-1), we have for binary CPM

$$\tilde{S}(t) = \sqrt{\frac{2E_b}{T_b}}\exp\left\{j\phi\left(t,\boldsymbol{\alpha}\right)\right\}, \quad nT_b \le t \le (n+1)T_b \tag{2.8-66}$$

For $h = 0.5$ partial-response CPM, where the frequency pulse has duration LT_b (remember from our previous discussion that the value of L used to approximate GMSK is a function of the value of BT_b of interest), Laurent showed after much manipulation that the complex envelope in (2.8-66) could be expressed as[20]

---

[19] The work was later extended to the M-ary case by Mengali and Morelli [53].

[20] For observation of the signal in the Nth transmission interval, $(N-1)T_b \le t \le NT_b$, the upper limit on n in (2.8-67) and (2.8-68) can be changed from ∞ to $N-1$ since the signal does not depend on future data bits. Furthermore, for a finite data sequence of length N, i.e., $\alpha_0, \alpha_1, \cdots, \alpha_{N-1}$, the lower limit on n in (2.8-67) and (2.8-68) can be changed from $-\infty$ to 0.

$$\tilde{S}(t) = \sqrt{\frac{2E_b}{T_b}} \sum_{K=0}^{2^{L-1}-1} \left[\sum_{n=-\infty}^{\infty} e^{j\frac{\pi}{2}A_{K,n}} C_K\left(t - nT_b\right) \right]$$

$$\triangleq \sqrt{\frac{2E_b}{T_b}} \sum_{K=0}^{2^{L-1}-1} \left[\sum_{n=-\infty}^{\infty} \tilde{a}_{K,n} C_K\left(t - nT_b\right) \right] \tag{2.8-67}$$

which results in the real CPM signal

$$s(t) = \sqrt{\frac{2E_b}{T_b}} \sum_{K=0}^{2^{L-1}-1} \left[\sum_{n=-\infty}^{\infty} C_K\left(t - nT_b\right) \cos\left(2\pi f_c t + \frac{\pi}{2} A_{K,n}\right) \right] \tag{2.8-68}$$

i.e., a superposition of 2^{L-1} amplitude-/phase-modulated pulse streams. In (2.8-68), $C_K(t)$ is the equivalent pulse shape for the kth AMP stream and is determined as follows:

First, define the generalized phase pulse function by

$$\Psi(t) = \begin{cases} \pi q(t), & 0 \le t \le LT_b \\ \dfrac{\pi}{2}\left[1 - 2q\left(t - LT_b\right)\right], & LT_b \le t \end{cases} \tag{2.8-69}$$

which is obtained by taking the nonconstant part of the phase pulse, $q(t)$, that exists in the interval $0 \le t \le LT_b$ and reflecting it about the $t = LT_b$ axis.[21] Therefore, in view of (2.8-69), $\Psi(t)$ is a waveform that is nonzero in the interval $0 \le t \le 2LT_b$ and symmetric around $t = LT_b$. The symmetry around $t = LT_b$ assumes that the frequency pulse, $g(t)$, is even symmetric around $t = LT_b/2$ and, thus, the phase pulse $q(t)$ is odd symmetric around the value $\pi/4$ at $t = LT_b/2$. Next define

$$\left.\begin{aligned} S_0(t) &\triangleq \sin\Psi(t) \\ S_n(t) &\triangleq S_0(t + nT) = \sin\Psi(t + nT) \end{aligned}\right\} \tag{2.8-70}$$

Finally,

[21] For the Laurent representation, it is convenient to shift the frequency pulse of (2.8-54) to the interval $0 \le t \le LT_b$ before integrating it to get the phase pulse, $q(t)$.

$$C_K(t) = S_0(t) \prod_{i=1}^{L-1} S_{i+L\beta_{K,i}}(t), \quad 0 \le K \le 2^{L-1} - 1, \quad 0 \le t \le T_{bK},$$

$$T_{bK} = T_b \times \min_{i=1,2,\cdots,L-1} \left[L\left(2 - \beta_{K,i}\right) - i \right] \quad (2.8\text{-}71)$$

where $\beta_{K,i}, i = 1, 2, \cdots, L-1$ are the coefficients in the binary representation of the integer K, i.e.,

$$K = \sum_{i=1}^{L-1} 2^{i-1} \beta_{K,i} \quad (2.8\text{-}72)$$

Note from (2.8-71) that each of the equivalent pulse waveforms, $C_K\,(t)$, in general have different durations, and, consequently, the pulse streams in (2.8-68) consist of overlapping pulses.

The complex phase coefficient $\tilde{a}_{K,n} \triangleq e^{j(\pi/2)A_{K,n}}$ associated with the nth T-s translate of this Kth pulse shape, namely $C_K(t - nT)$, is also expressible in terms of the binary representation of the integer K as given in (2.8-72). In particular,

$$\left.\begin{aligned} A_{K,n} &= \sum_{i=-\infty}^{n} \alpha_i - \sum_{i=1}^{L-1} \alpha_{n-i}\beta_{K,i} = A_{0,n} - \sum_{i=1}^{L-1} \alpha_{n-i}\beta_{K,i} \\ A_{0,n} &= \alpha_n + A_{0,n-1} \end{aligned}\right\} \quad (2.8\text{-}73)$$

and thus,

$$\tilde{a}_{K,n} \triangleq e^{j(\pi/2)A_{K,n}} = \exp\left[j\frac{\pi}{2} \left(A_{0,n-L} + \sum_{i=0}^{L-1} \alpha_{n-i} - \sum_{i=1}^{L-1} \alpha_{n-i}\beta_{K,i} \right) \right]$$

$$= \tilde{a}_{0,n-L} e^{j(\pi/2)\alpha_n} \prod_{i=1}^{L-1} e^{j(\pi/2)\alpha_{n-i}[1-\beta_{K,i}]} \quad (2.8\text{-}74)$$

Before proceeding further, we present an example corresponding to a particular value of L to illustrate the above description of the representation. Consider the case of $L = 4$, which as previously mentioned, is adequate to represent GMSK with $BT_b \ge 0.25$. Therefore, from (2.8-71), there are $2^{L-1} = 8$ different $C_K\,(t)$'s,

i.e., $C_0(t), C_1(t), \cdots, C_7(t)$, each of which is a product of the basic generalized pulse shape $S_0(t)$ and $L-1=3$ other $S_i(t)$'s with the particular ones being chosen according to the coefficients in the binary representation of the index, K. For example, for $K=3$, we would have

$$K = 3 = 2^0 \times 1 + 2^1 \times 1 + 2^2 \times 0 \quad \Rightarrow \quad \beta_{3,1} = 1, \beta_{3,2} = 1, \beta_{3,3} = 0 \tag{2.8-75}$$

and thus,

$$C_3(t) = S_0(t) \prod_{i=1}^{3} S_{i+4\beta_{3,i}}(t) = S_0(t)S_5(t)S_6(t)S_3(t), \quad 0 \le t \le T_{b3} = 2T_b \tag{2.8-76}$$

In summary,

$$\left.\begin{aligned}
C_0(t) &= S_0(t)S_1(t)S_2(t)S_3(t), \quad 0 \le t \le 5T_b \\
C_1(t) &= S_0(t)S_2(t)S_3(t)S_5(t), \quad 0 \le t \le 3T_b \\
C_2(t) &= S_0(t)S_1(t)S_3(t)S_6(t), \quad 0 \le t \le 2T_b \\
C_3(t) &= S_0(t)S_3(t)S_5(t)S_6(t), \quad 0 \le t \le 2T_b \\
C_4(t) &= S_0(t)S_1(t)S_2(t)S_7(t), \quad 0 \le t \le T_b \\
C_5(t) &= S_0(t)S_2(t)S_5(t)S_7(t), \quad 0 \le t \le T_b \\
C_6(t) &= S_0(t)S_1(t)S_6(t)S_7(t), \quad 0 \le t \le T_b \\
C_7(t) &= S_0(t)S_5(t)S_6(t)S_7(t), \quad 0 \le t \le T_b
\end{aligned}\right\} \tag{2.8-77}$$

From (2.8-74) the set of complex phase coefficients for the third pulse train corresponding to $C_3(t)$ of (2.8-76) would be

$$\tilde{a}_{3,n} = \tilde{a}_{0,n-4} e^{j(\pi/2)\alpha_n} \prod_{i=1}^{3} e^{j(\pi/2)\alpha_{n-i}[1-\beta_{3,i}]} = \tilde{a}_{0,n-4} e^{j(\pi/2)\alpha_n} e^{j(\pi/2)\alpha_{n-3}} \tag{2.8-78}$$

The complete group of phase coefficient sets for all eight pulse trains is given by (also see [46, Eq. (A.19)])

$$\left.\begin{aligned}
\tilde{a}_{0,n} &= \tilde{a}_{0,n-4}e^{j(\pi/2)\alpha_n}e^{j(\pi/2)\alpha_{n-1}}e^{j(\pi/2)\alpha_{n-2}}e^{j(\pi/2)\alpha_{n-3}} \\
&= j^{\alpha_n+\alpha_{n-1}+\alpha_{n-2}+\alpha_{n-3}}\tilde{a}_{0,n-4} = j^{\alpha_n}\tilde{a}_{0,n-1} \\
\tilde{a}_{1,n} &= \tilde{a}_{0,n-4}e^{j(\pi/2)\alpha_n}e^{j(\pi/2)\alpha_{n-2}}e^{j(\pi/2)\alpha_{n-3}} \\
&= j^{\alpha_n+\alpha_{n-2}+\alpha_{n-3}}\tilde{a}_{0,n-4} = j^{\alpha_n}\tilde{a}_{0,n-2} \\
\tilde{a}_{2,n} &= \tilde{a}_{0,n-4}e^{j(\pi/2)\alpha_n}e^{j(\pi/2)\alpha_{n-1}}e^{j(\pi/2)\alpha_{n-3}} \\
&= j^{\alpha_n+\alpha_{n-1}+\alpha_{n-3}}\tilde{a}_{0,n-4} = j^{\alpha_n+\alpha_{n-1}}\tilde{a}_{0,n-3} \\
\tilde{a}_{3,n} &= \tilde{a}_{0,n-4}e^{j(\pi/2)\alpha_n}e^{j(\pi/2)\alpha_{n-3}} = j^{\alpha_n+\alpha_{n-3}}\tilde{a}_{0,n-4} = j^{\alpha_n}\tilde{a}_{0,n-3} \\
\tilde{a}_{4,n} &= \tilde{a}_{0,n-4}e^{j(\pi/2)\alpha_n}e^{j(\pi/2)\alpha_{n-1}}e^{j(\pi/2)\alpha_{n-2}} = j^{\alpha_n+\alpha_{n-1}+\alpha_{n-2}}\tilde{a}_{0,n-4} \\
\tilde{a}_{5,n} &= \tilde{a}_{0,n-4}e^{j(\pi/2)\alpha_n}e^{j(\pi/2)\alpha_{n-2}} = j^{\alpha_n+\alpha_{n-2}}\tilde{a}_{0,n-4} \\
\tilde{a}_{6,n} &= \tilde{a}_{0,n-4}e^{j(\pi/2)\alpha_n}e^{j(\pi/2)\alpha_{n-1}} = j^{\alpha_n+\alpha_{n-1}}\tilde{a}_{0,n-4} \\
\tilde{a}_{7,n} &= \tilde{a}_{0,n-4}e^{j(\pi/2)\alpha_n} = j^{\alpha_n}\tilde{a}_{0,n-4}
\end{aligned}\right\} \quad (2.8\text{-}79)$$

It is to be emphasized that to the extent that GMSK can be approximated by a partial-response CPM with finite L, the AMP representation is *exact*. For the case of $L = 4$, Ref. 46 states that, based on computer simulations, the first AMP component corresponding to the pulse stream $\{C_0(t - nT)\}$ contains the fraction 0.991944 of the total signal energy, and the second component corresponding to the pulse stream $\{C_1(t - nT)\}$ contains the fraction 0.00803 of the total energy. Thus, the remaining six components contain only the fraction 2.63×10^{-5} of the total signal energy, and thus, to a good approximation can be ignored. Hence, we conclude that for values of BT_b, where $L = 4$ is an appropriate value for the truncation of the frequency pulse, *a two pulse stream AMP representation corresponding to $K = 1$ and $K = 2$ is sufficient to approximate GMSK*, i.e.,

$$\tilde{S}_{\text{GMSK}}(t) = \sqrt{\frac{2E_b}{T_b}}\left[\sum_{n=-\infty}^{\infty}\tilde{a}_{0,n}C_0\left(t - nT_b\right) + \sum_{n=-\infty}^{\infty}\tilde{a}_{1,n}C_1\left(t - nT_b\right)\right] \quad (2.8\text{-}80)$$

where $C_0\left(t\right)$ and $C_1\left(t\right)$ are determined from the first two equations in (2.8-77) (see Fig. 2-32 for an illustration of these two waveforms) and, likewise, $\tilde{a}_{0,n}$ and $\tilde{a}_{1,n}$ are determined from the first two equations of (2.8-79). Since the actual data symbols, $\{\alpha_n\}$, range over the values ± 1, then the even and odd complex symbols for each of the two pulse streams take on values

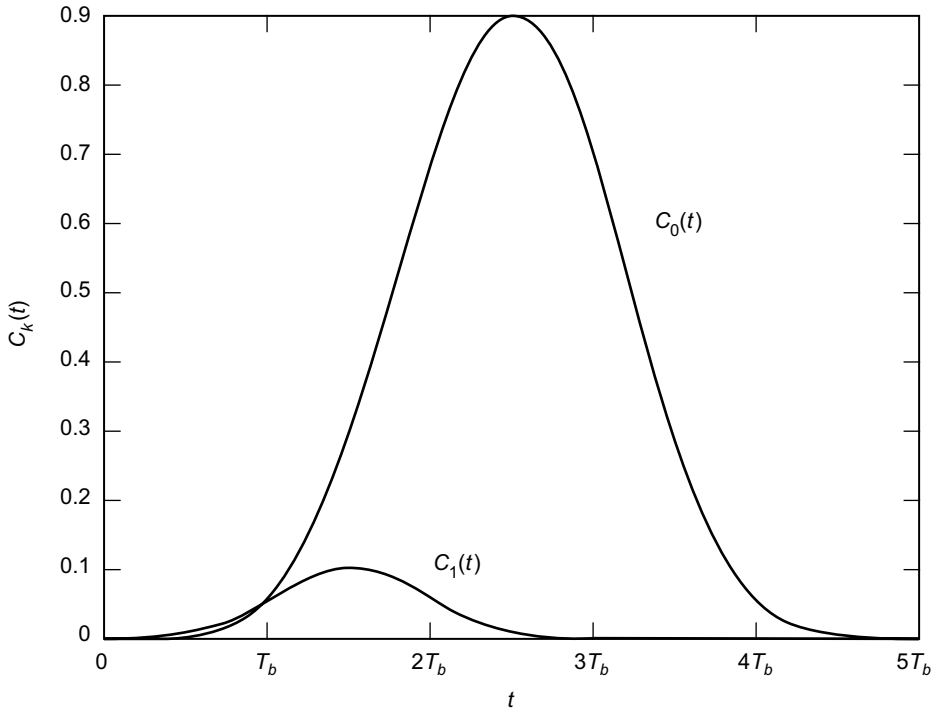

Fig. 2-32. Pulse shapes for first and second AMP streams.

$$\left.\begin{aligned} \{\tilde{a}_{0,2n}\} \in \{j,-j\}\,, \quad \{\tilde{a}_{0,2n+1}\} \in \{1,-1\} \\ \{\tilde{a}_{1,2n}\} \in \{1,-1\}\,, \quad \{\tilde{a}_{1,2n+1}\} \in \{j,-j\} \end{aligned}\right\} \tag{2.8-81}$$

which clearly indicate a representation composed of the superposition of two I-Q signals. Note that the sequences $\{\tilde{a}_{0,n}\}$ and $\{\tilde{a}_{1,n}\}$ are themselves uncorrelated as well as being mutually uncorrelated, viz.,

$$\begin{aligned} E\left\{\tilde{a}_{0,k}\tilde{a}^*_{1,k-m}\right\} &= E\left\{j\alpha_k j\alpha_{k-1}\cdots j\alpha_{k-m-1}\tilde{a}_{0,k-m-2} \times -j\alpha_{k-m}\tilde{a}^*_{0,k-m-2}\right\} \\ &= \pm E\left\{\alpha_k\alpha_{k-1}\cdots\alpha_{k-m-1}\alpha_{k-m}\right\} = 0, \quad m > 0 \end{aligned} \tag{2.8-82}$$

Furthermore, since for binary ± 1 data, $j^{\alpha_n} = j\alpha_n$, then the first two equations of (2.8-79) become

$$\left.\begin{aligned}\tilde{a}_{0,n} &= j\alpha_n \tilde{a}_{0,n-1} \\ \tilde{a}_{1,n} &= j\alpha_n \tilde{a}_{0,n-2}\end{aligned}\right\} \tag{2.8-83}$$

which clearly identifies the fact that the complex symbols for the two pulse streams are obtained from a *differentially encoded* version of the input data. Finally, the corresponding real (± 1) symbols on the I and Q channels for the two pulse streams are

$$\left.\begin{aligned}a_{0,n}^{I} &= \tilde{a}_{0,2n+1} = \mathrm{Re}\left\{\tilde{a}_{0,2n+1}\right\} \\ a_{0,n}^{Q} &= -j\tilde{a}_{0,2n} = \mathrm{Im}\left\{\tilde{a}_{0,2n}\right\} \\ a_{1,n}^{Q} &= -j\tilde{a}_{1,2n+1} = \mathrm{Im}\left\{\tilde{a}_{1,2n+1}\right\} \\ a_{1,n}^{I} &= \tilde{a}_{1,2n} = \mathrm{Re}\left\{\tilde{a}_{1,2n}\right\}\end{aligned}\right\} \tag{2.8-84}$$

and, hence, the real GMSK two pulse stream approximation corresponding to (2.8-80) is

$$\begin{aligned}s_{\mathrm{GMSK}}(t) = \sqrt{\frac{2E_b}{T_b}}\Bigg[&\sum_{n=-\infty}^{\infty} a_{0,n}^{I} C_0\left(t-(2n+1)T_b\right)\cos 2\pi f_c t \\ &- \sum_{n=-\infty}^{\infty} a_{0,n}^{Q} C_0\left(t-2nT_b\right)\sin 2\pi f_c t \\ &+ \sum_{n=-\infty}^{\infty} a_{1,n}^{I} C_1\left(t-2nT_b\right)\cos 2\pi f_c t \\ &- \sum_{n=-\infty}^{\infty} a_{1,n}^{Q} C_1\left(t-(2n+1)T_b\right)\sin 2\pi f_c t\Bigg]\end{aligned} \tag{2.8-85}$$

which has the implementation of Fig. 2-33. Note that each of the pulse streams is in the form of a pulse-shaped OQPSK modulation with overlapping pulses and effective symbol rate $T_s = 2T_b$ on each of the quadrature channels. Also, the encoding of the first pulse stream is a conventional differential encoder whereas the second pulse stream is generated from a product of the input data stream and a delayed version of the differentially encoded output of the first stream. Therefore, from a data encoding standpoint, the first pulse stream resembles MSK whereas the second does not.

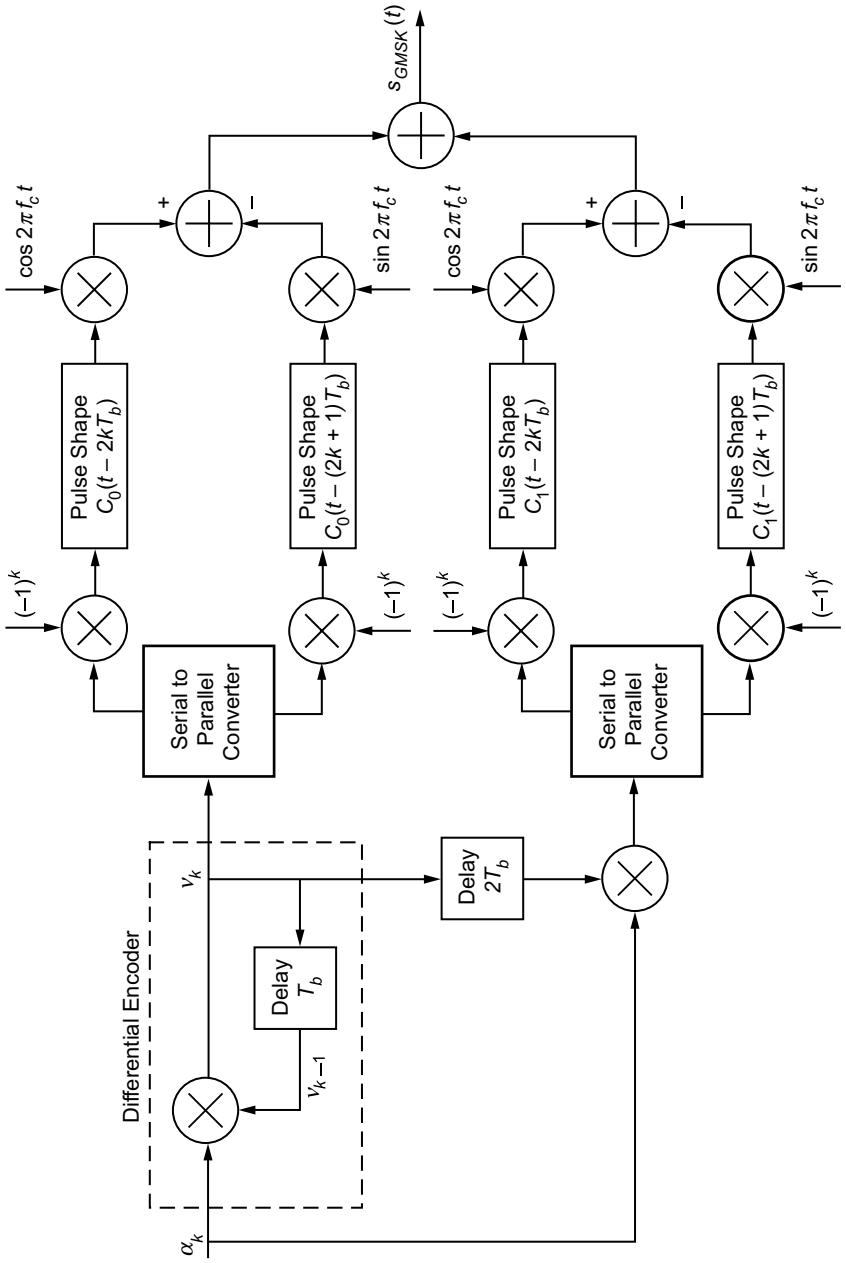

Fig. 2-33. Two-pulse stream I-Q implementation of GMSK.

b. Precoded GMSK. Because of the different encodings for the first and second pulse streams, we can only fully compensate for one of the two with a precoding operation. Thus, proceeding as in the MSK case, if we precede the GMSK modulator with a differential decoder [see Fig. 2-34(a)], then, as was true for MSK, the first pulse stream of the equivalent I-Q representation would no longer have a differential encoder at its front end. The effect of the precoding on the second pulse stream of the equivalent I-Q representation is to produce a particular feed-forward type of encoding [see Fig. 2-34(b)] that can be shown to be equivalent to a two-stage differential decoder (see Fig. 2-35). Such precoded GMSK has been considered by several authors in the literature [54–57] as a means of improving receiver performance.

2.8.2.4 Power Spectral Density Considerations. As mentioned above, one advantage of the Laurent representation is that it provides a simple means of computing the PSD. In particular, since the various pulse-stream equivalent data sequences are both self- and mutually uncorrelated, for the GMSK signal with complex form as in (2.8-67), the PSD is simply

$$S(f) = E_b \sum_{k=0}^{2^{L-1}-1} \frac{1}{T_b} |\mathcal{C}_k(f)|^2, \quad \mathcal{C}_k(f) = \mathcal{F}\{C_k(t)\} \tag{2.8-86}$$

or for $L = 4$ and the two pulse stream approximation of (2.8-80),

$$S(f) = \frac{E_b}{T_b} \left[|\mathcal{C}_0(f)|^2 + |\mathcal{C}_1(f)|^2 \right] \tag{2.8-87}$$

Figure 2-36 is a plot of the normalized (all curves start at zero decibels at zero frequency) GMSK PSD as computed from (2.8-86), with frequency pulse length in bits, L, as a parameter. The values of BT_b have been chosen equal to the reciprocal of L. Thus, for example, a value of $L = 4$ results in a curve for $BT_b = 0.25$ that corresponds to the special case previously considered. Comparing

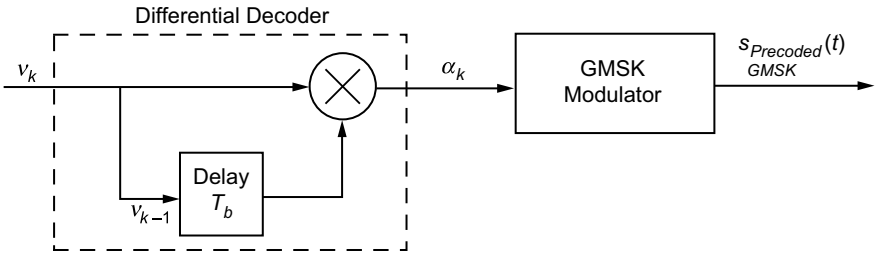

Fig. 2-34(a). Precoded GMSK transmitter.

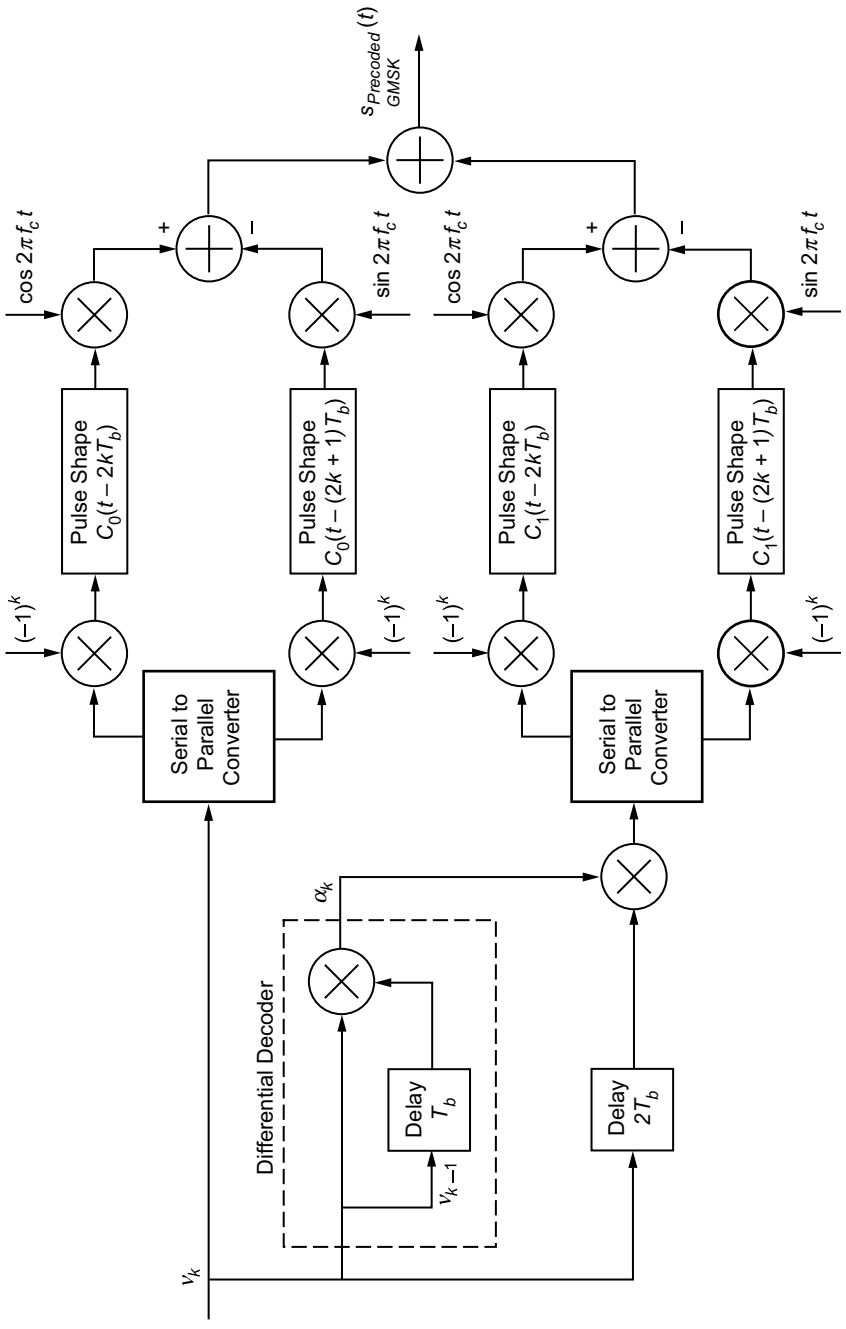

Fig. 2-34(b). Two-pulse stream I-Q implementation of precoded GMSK.

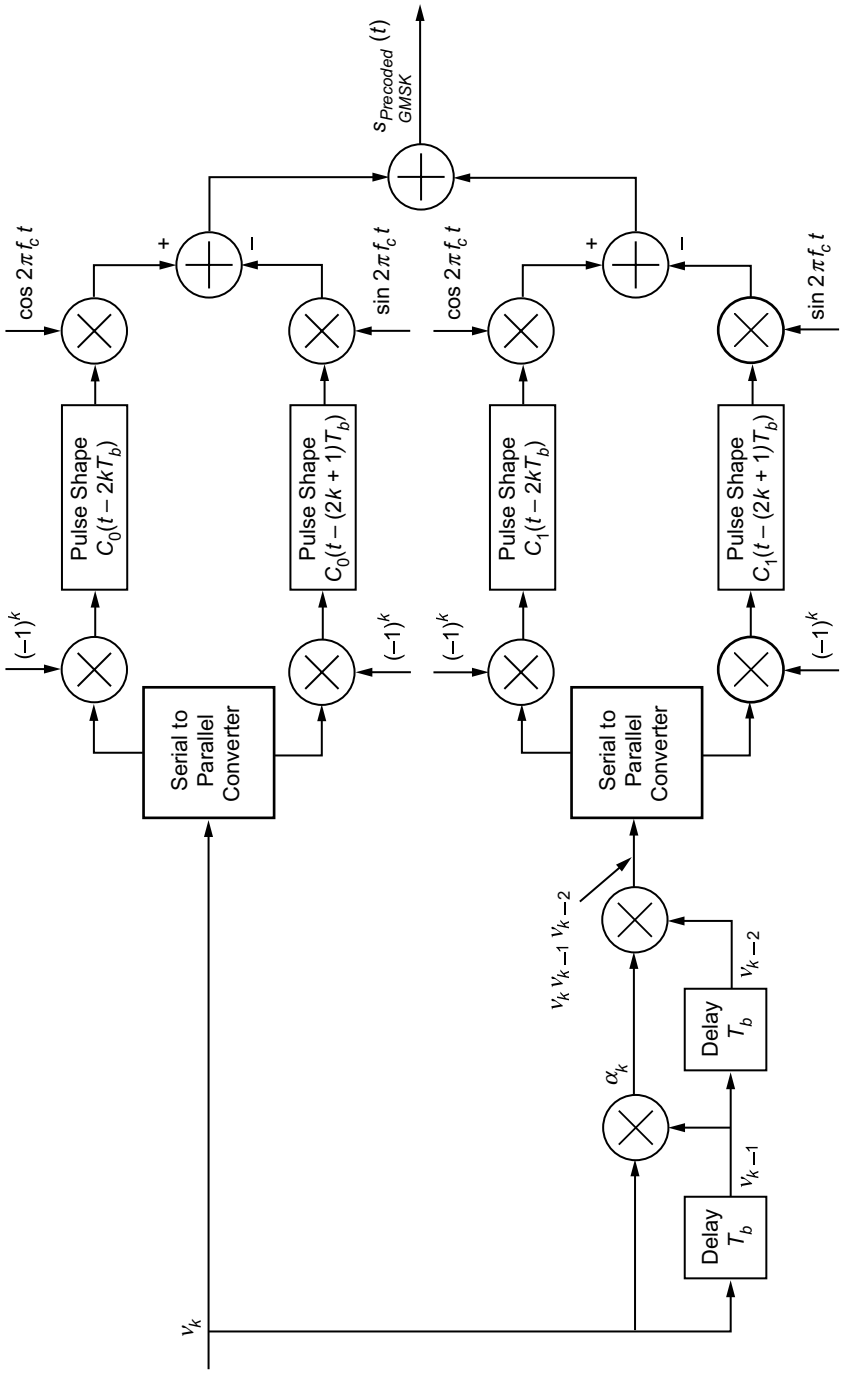

Fig. 2-35. Alternative two-pulse stream I-Q implementation of precoded GMSK.

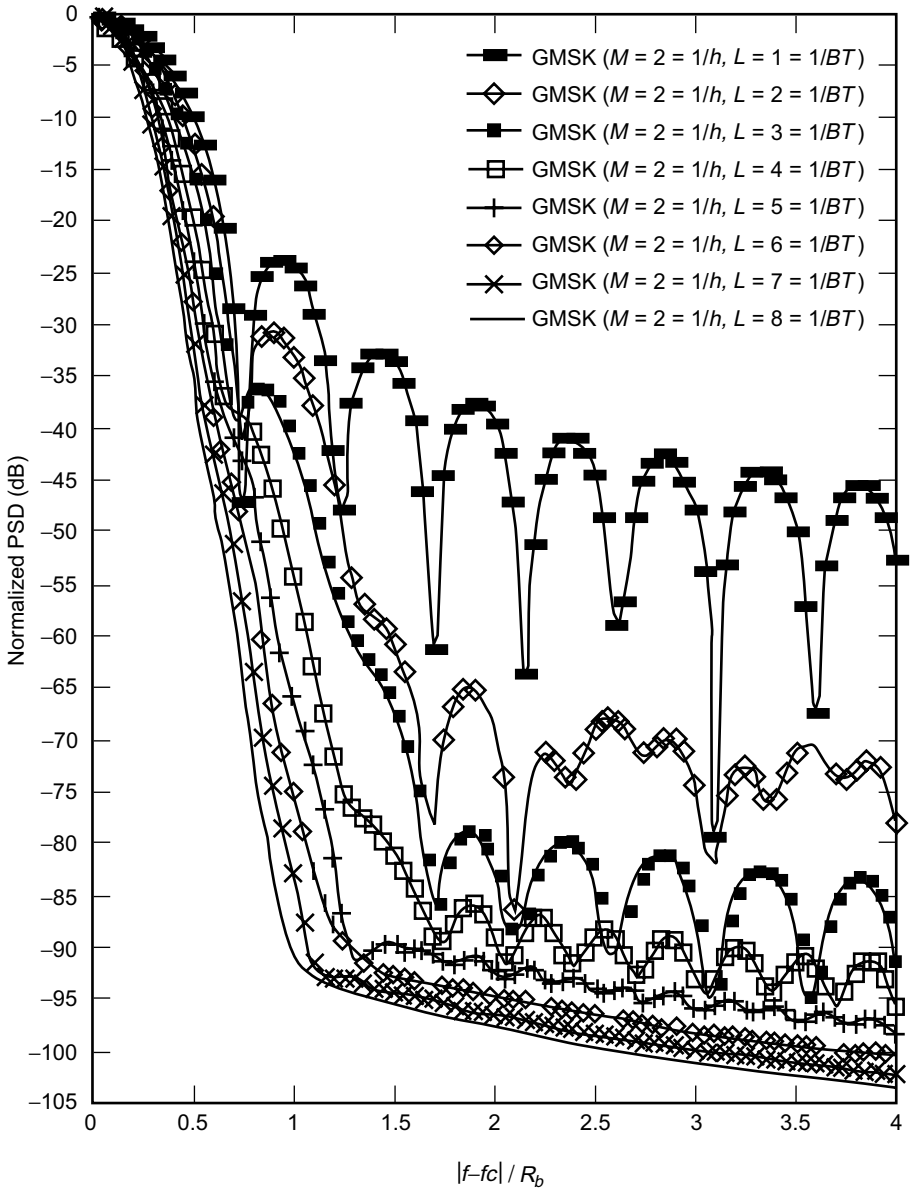

Fig. 2-36. Power spectral density of GMSK with BT_b as a parameter. Redrawn from [54].

Fig. 2-36 with Fig. 2-15, we observe the potentially significant improvement in spectral efficiency of the partial-response CPM modulation (GMSK) over the full-response CPM modulation (MSK), when the value of BT_b is sufficiently less than unity. Finally, we note that the PSD of precoded GMSK is identical to that of GMSK.

2.8.2.5 Approximate AMP Representation of GMSK Based on a Single Pulse Stream. Before moving on to a discussion of the various types of receivers that have been proposed for GMSK, it is instructive to further approximate (simplify) the AMP representation, since the structure of one of these receivers stems from this approximation. In the AMP representation of (2.8-67) or (2.8-68), the dominant term is the pulse stream corresponding to $C_0(t)$ (for full-response CPM, i.e., $L = 1$, it would be the only one), since its duration is the longest (at least $2T_b$ longer than any other pulse component) and it also conveys the most significant part of the total energy of the signal. (Although Laurent never proves the latter mathematically, it appears to be the case for all practical CPM scenarios.) Thus, approximating the AMP representation with a single pulse stream, which is an exact representation for MSK, is a reasonable thing to do. As such, we propose an approximation of (2.8-67), where the pulse shape, $P(t)$ (called the "main pulse" in Ref. 52), used for the single-stream AMP representation should have the same phase shift as that associated with $C_0(t)$ and must satisfy some optimization criterion in the sense of being the best approximation of the signal, viz.,

$$\hat{\tilde{S}}(t) = \sum_{n=-\infty}^{\infty} e^{j(\pi/2)A_{0,n}} P(t - nT_b) = \sum_{n=-\infty}^{\infty} e^{j(\pi/2)\sum_{i=0}^{n}\alpha_i} P(t - nT_b) \quad (2.8\text{-}88)$$

The optimization criterion selected by Laurent consisted of minimizing the average energy of the difference between the complete signal and its approximation. Two methods are proposed in Ref. 52 for solving this optimization problem in the general case of CPM with modulation index, h, the second of which is preferred because it illustrates the important properties of the main pulse. In particular, $P(t)$ is expressed as a weighted superposition of time-shifted versions of the finite duration components, $C_i(t)$. It is further shown in Ref. 52, that regardless of the value of L, for $h = 0.5$ (as is the case for GMSK), the main pulse is simply given by $C_0(t)$. Hence, we conclude that using only the first AMP component of the signal is the best—and naturally the simplest—possible approximation in the above mean-square energy sense.

2.8.2.6 Coherent GMSK Receivers and Their Performance. A variety of different types of receivers [46,54,56,57] have been proposed for coherent

detection of GMSK; most of them are based on the Laurent representation and employ the Viterbi Algorithm (VA) [58]. To start the discussion, we consider first the optimum receiver based upon an L-bit duration GMSK frequency pulse.

Because of the memory inherent in CPM, regardless of its mathematical representation, the optimum receiver (from the standpoint of minimizing the message error probability) has the form of an MLSE which is typically implemented using the VA. This receiver employs $m = 2^{L-1} - 1$ filters matched to each of the m pulse shapes in the complex baseband AMP representation of (2.8-67). These filters act on the received complex signal plus noise, and their outputs are inputted to a VA whose decision metric is based upon the equivalent data stream encodings of (2.8-79) (see Fig. 2-37). The number of states in the trellis diagram characterizing the receiver is equal to 2^L, e.g., $L = 4$ would require a 16-state trellis.

a. Optimum Receiver. When the GMSK signal of (2.8-68) is transmitted over an AWGN channel, the received signal is given by

$$z(t) = s(t) + n(t) \tag{2.8-89}$$

where $n(t)$ is as before a zero mean Gaussian process, independent of the signal, with single-sided PSD equal to N_0 watts/hertz. Since for a length N data

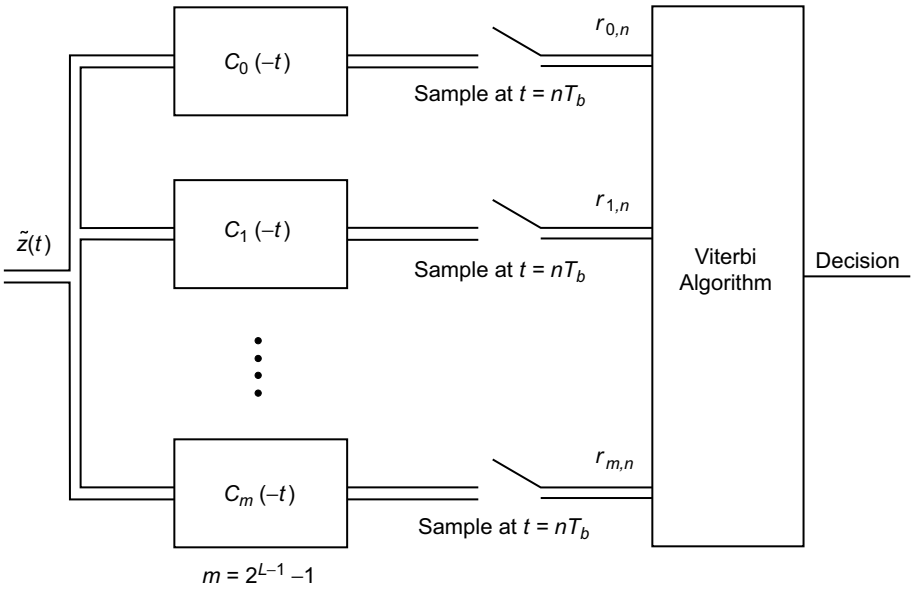

Fig. 2-37. Optimum GMSK receiver.

sequence, all of the possible 2^N transmitted signals have equal energy and are equally likely, the optimum receiver that minimizes the message (sequence) error probability chooses that message, i, that maximizes the metric

$$\Lambda_i = 2\int_{-\infty}^{\infty} z(t)s_i(t)dt = \mathrm{Re}\left\{\int_{-\infty}^{\infty} \tilde{Z}(t)\tilde{S}_i^*(t)dt\right\} \tag{2.8-90}$$

where the second equality ignores the second harmonic carrier term, $s_i(t)$ is the signal corresponding to the ith data sequence with complex envelope $\tilde{S}_i(t)$, and analogous to (2.8-65)

$$z(t) = \mathrm{Re}\left\{\tilde{Z}(t)e^{j2\pi f_c t}\right\} \tag{2.8-91}$$

Substituting (2.8-67) in (2.8-90) yields an additive form for the metric, namely,

$$\Lambda_i = \sqrt{\frac{2E_b}{T_b}}\sum_{n=0}^{N-1}\lambda_i(n) \tag{2.8-92}$$

where $\lambda_i(n)$ is the trellis branch metric given by

$$\lambda_i(n) = \mathrm{Re}\left\{\sum_{K=0}^{2^{L-1}-1} {\tilde{a}_{K,n}^i}^* \int_{-\infty}^{\infty} \tilde{Z}(t)C_K\left(t-nT_b\right)dt\right\}$$

$$\triangleq \mathrm{Re}\left\{\sum_{K=0}^{2^{L-1}-1} {\tilde{a}_{K,n}^i}^* r_{K,n}\right\} \tag{2.8-93}$$

The superscript "i" on the equivalent complex data symbols denotes the ith data sequence, i.e., these are the N symbols corresponding to the signal, $s_i(t)$. The correlation values

$$r_{K,n} = \int_{-\infty}^{\infty} \tilde{Z}(t)C_K\left(t-nT_b\right)dt, \quad 0 \le K \le 2^{L-1}-1, \quad 0 \le n \le N \tag{2.8-94}$$

are thus sufficient statistics for making the message decision and can be obtained for any fixed n by sampling the outputs of the bank of 2^{L-1} matched filters in Fig. 2-37 at time $t = nT_b$.

Computation of an nth branch metric requires knowledge of the equivalent complex data sequence $\{\tilde{a}_{K,n}\}$. This in turn can be found from the current data symbol α_n and a state defined by the vector $(a_{0,n-L}, \alpha_{n-L+1}, \alpha_{n-L+2}, \cdots, \alpha_{n-2}, \alpha_{n-1})$. Therefore, the decision rule can be implemented by inputting the set of matched filter output samples to a VA using the above state definition and current symbol to define the trellis states and transitions between them. The complexity of the VA is proportional to the number of states, which as previously mentioned, is equal to 2^L.

b. Simplified (Suboptimum) GMSK Receivers. Using the approximate AMP representation discussed in Sec. 2.8.2.5, Kaleh [46] first derived a reduced-complexity Viterbi detector that achieved near-optimal performance. By "reduced-complexity," we mean that the number of matched filters and VA states is appreciably smaller than would be required for the truly optimum receiver. In particular, a receiver consisting of only two matched filters and a four-state VA resulted in a performance degradation of less than 0.24 dB relative to the optimum and much more complex receiver. In addition to the simplification of the optimum receiver based on an error probability criterion, Kaleh also considered an optimum coherent linear receiver based on a minimum mean-square error (MMSE) criterion. Such a receiver has a generic form analogous to that used for detection of MSK (as such it was referred to in Ref. 46 as an MSK-type receiver) except for the fact that the receive filter is composed now of a combination of matched and Wiener-type filters. In what follows, we explore these two receiver options.

The complexity of the optimum MLSE receiver can be reduced by approximating the AMP representation with a smaller number of pulse streams, as previously discussed. In particular, consider replacing the 2^{L-1} pulse streams in (2.8-67) by the first $\hat{K}$ of them, where $\hat{K}$ is chosen such that the remaining components cumulatively have very small energy [(2.8-80) is a particular example of this where $\hat{K} = 2$]. As such, we can write the transmitted signal in the approximate (simplified) complex baseband form

$$\hat{\tilde{S}}(t) = \sqrt{\frac{2E_b}{T_b}} \sum_{K=0}^{\hat{K}-1} \left[\sum_{n=-\infty}^{\infty} e^{j\pi h A_{K,n}} C_K (t - nT_b) \right]$$

$$\triangleq \sqrt{\frac{2E_b}{T_b}} \sum_{K=0}^{\hat{K}-1} \left[\sum_{n=-\infty}^{\infty} \tilde{a}_{K,n} C_K (t - nT_b) \right] \tag{2.8-95}$$

where the "hat" is used to denote approximation. Then, in accorandance with (2.8-92) the simplified receiver computes the approximate metric

$$\hat{\Lambda}_i = \sqrt{\frac{2E_b}{T_b}} \sum_{n=0}^{N-1} \hat{\lambda}_i(n) \tag{2.8-96}$$

where

$$\hat{\lambda}_i(n) = \mathrm{Re}\left\{ \sum_{K=0}^{\hat{K}-1} \tilde{a}^i_{K,n}{}^* \int_{-\infty}^{\infty} \hat{\tilde{Z}}(t) C_K\left(t - nT_b\right) dt \right\} \triangleq \mathrm{Re}\left\{ \sum_{K=0}^{\hat{K}-1} \tilde{a}^i_{K,n}{}^* r_{K,n} \right\} \tag{2.8-97}$$

and $\hat{\tilde{Z}}(t)$ corresponds to the received version of $\hat{\tilde{S}}(t)$. Since $r_{K,n}, K = \hat{K}, \hat{K} + 1, \cdots, 2^{L-1} - 1$ are considered as irrelevant, the number of matched filters needed in Fig. 2-37 is reduced from 2^{L-1} to $\hat{K}$. Also, a great reduction in the complexity of the VA is achieved, since the number of states can accordingly be reduced from 2^L to $2\hat{K}$.

Pursuing now in detail the case where $\hat{K} = 2$, the even branch metrics in (2.8-97) are given by

$$\begin{aligned}
\hat{\lambda}_i(2n) &= \mathrm{Re}\left\{ \tilde{a}^i_{0,2n}{}^* r_{0,2n} + \tilde{a}^i_{1,2n}{}^* r_{1,2n} \right\} \\
&= \mathrm{Re}\left\{ a^i_{0,2n}{}^* r_{0,2n} \right\} + \mathrm{Re}\left\{ -j\alpha^i_{2n} \tilde{a}^i_{0,2n-2}{}^* r_{1,2n} \right\} \\
&= a^i_{0,2n} \mathrm{Re}\left\{ r_{0,2n} \right\} + \mathrm{Re}\left\{ -\frac{\tilde{a}^i_{0,2n}}{\tilde{a}^i_{0,2n-1}} \tilde{a}^i_{0,2n-2}{}^* r_{1,2n} \right\} \\
&= a^i_{0,2n} \mathrm{Re}\left\{ r_{0,2n} \right\} + \mathrm{Re}\left\{ -a^i_{0,2n} j a^i_{0,2n-1} a^i_{0,2n-2} r_{1,2n} \right\} \\
&= \overbrace{a^i_{0,2n}}^{\text{current bit}} \mathrm{Re}\left\{ r_{0,2n} \right\} + \overbrace{a^i_{0,2n}}^{\text{current bit}} \overbrace{a^i_{0,2n-1} a^i_{0,2n-2}}^{\text{2 previous bits}} \mathrm{Im}\left\{ r_{1,2n} \right\}
\end{aligned} \tag{2.8-98a}$$

whereas the odd branch metrics are given by

$$\hat{\lambda}_i(2n-1) = \overbrace{a^i_{0,2n-1}}^{\text{current bit}} \mathrm{Im}\left\{ r_{0,2n-1} \right\} - \overbrace{a^i_{0,2n-1}}^{\text{current bit}} \overbrace{a^i_{0,2n-2} a^i_{0,2n-3}}^{\text{2 previous bits}} \mathrm{Re}\left\{ r_{1,2n-1} \right\} \tag{2.8-98b}$$

Consequently, at any time, nT_b, we see that the state vector is defined by $a^i_{0,n-1}a^i_{0,n-2}$, i.e., the two previous equivalent real bits, which results in the four-state trellis illustrated in Fig. 2-38. The VA makes decisions, $\hat{a}_n$, on the real equivalent bits, a_n, from which the decisions on the actual transmitted bits are obtained from the differential decoding operation

$$\left.\begin{aligned} \hat{\alpha}_{2n} &= -\hat{a}_{2n}\hat{a}_{2n-1} \\ \hat{\alpha}_{2n+1} &= \hat{a}_{2n+1}\hat{a}_{2n} \end{aligned}\right\} \tag{2.8-99}$$

The performance of the simplified Viterbi receiver was computed in Ref. 46, based on an upper bound obtained from the minimum Euclidean distance of the signaling set. In particular, it is well known that for modulations characterized by a trellis-type decoding algorithm, the bit error probability is upper bounded by

$$P_b(E) \leq CQ\left(\frac{d_{\min}}{\sqrt{2N_0}}\right) \tag{2.8-100}$$

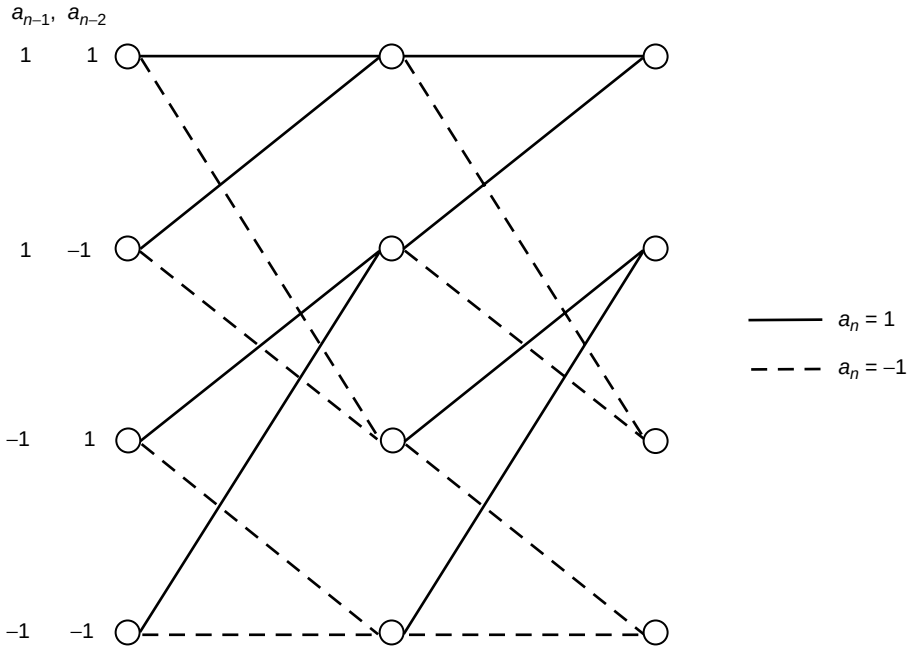

Fig. 2-38. Trellis diagram for a simplified Viterbi receiver for GMSK; $BT_b = 0.25$.

where C is a constant that depends on the number of nearest neighbors in the constellation to the transmitted signal and $d_{\min}$ is the minimum Euclidean distance between transmitted signals. Using this measure of performance, it was shown in Ref. 46 that *the simplified Viterbi receiver that uses two matched filters and a 4-state VA has a degradation of less than 0.24 dB when compared to the optimum Viterbi receiver that requires 8 matched filters and a 16-state VA.*

Next, we consider the implementation of simple MSK-type linear receivers for GMSK (see Fig. 2-39). Such receivers are memoryless and make bit-by-bit decisions on the transmitted data. In the case of true MSK, the receiver operates in the absence of ISI and, thus, the receive filter is merely the matched filter to the transmitted amplitude pulse shape, i.e., $C_0(t) = S_0(t) = \sin \pi t/2T_b$. Even in the case of generalized MSK with $h = 0.5$, the receive filter still operates in the absence of ISI with a matched receive filter in accordance with $C_0(t)$, which from (2.8-69) and (2.8-70) would be now given by

$$C_0(t) = S_0(t) = \begin{cases} \sin\big(\pi q(t)\big), & 0 \le t \le T_b \\ \sin\left(\dfrac{\pi}{2}\big[1 - 2q(t - T_b)\big]\right), & T_b \le t \le 2T_b \end{cases} \tag{2.8-101}$$

Before showing how such a receiver must be modified for GMSK, we first review its application to MSK.

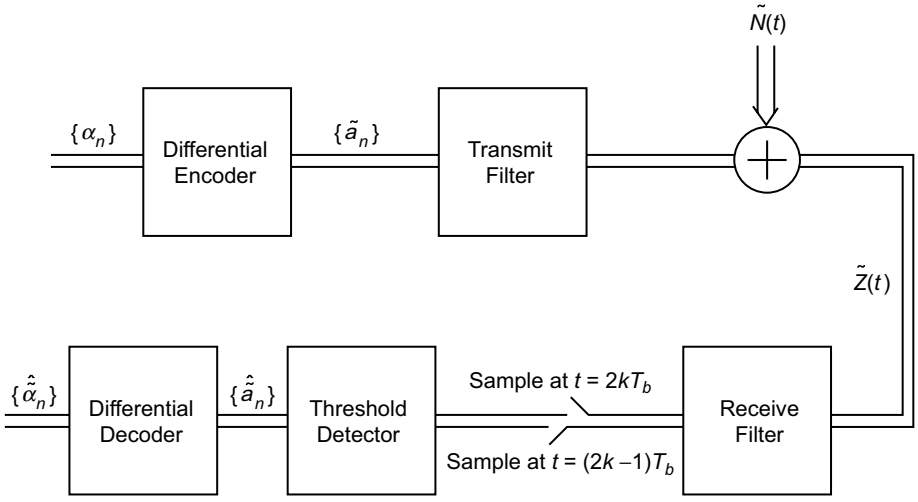

Fig. 2-39. Baseband model of an MSK-type system for GMSK.

For MSK with $L = 1$, the branch metric of (2.8-93) would simplify to

$$\lambda_i(n) = \mathrm{Re}\left\{\tilde{a}_{0,n}^{i\ *} r_{0,n}\right\} = \mathrm{Re}\left\{\tilde{a}_{0,n}^{i}\right\}\mathrm{Re}\left\{r_{0,n}\right\} + \mathrm{Im}\left\{\tilde{a}_{0,n}^{i}\right\}\mathrm{Im}\left\{r_{0,n}\right\}$$

$$= \begin{cases} a_{0,n}\mathrm{Re}\left\{r_{0,n}\right\}, & n \text{ even} \\ a_{0,n}\mathrm{Im}\left\{r_{0,n}\right\}, & n \text{ odd} \end{cases} \tag{2.8-102}$$

where

$$r_{0,n} = \int_{-\infty}^{\infty} \tilde{Z}(t) C_0\left(t - nT_b\right) dt, \quad 0 \le n \le N \tag{2.8-103}$$

which is the output of a single filter matched to $C_0(t)$ and sampled at time $t = nT_b$. Since the branch metric only depends on the current symbol, a memoryless receiver is appropriate for making decisions on the equivalent real data bits, $\{a_{0,n}\}$. Therefore, if we alternately sample the real and imaginary parts of the matched filter output at intervals of T_b s, we obtain ISI-free decisions on these bits. The true data bits are still obtained by following these decisions with the differential decoding operation of (2.8-99).

For GMSK, a superimposed I-Q representation is still possible. However, because the equivalent pulse shapes now spread over many symbol intervals and because more than one AMP component is present, the receive filter in Fig. 2-39 must be more complex than just a simple matched filter in order to account for the ISI inherent in the signal and the interference produced by the other AMP component(s). The nature of the modification of the receive filter required to accommodate these additional degradations is discussed below.

Consider then the transmitted signal of (2.8-95), where we explicitly substitute $\hat{K} = 2$ so as to correspond to the approximation discussed above for GMSK. Omitting herein the "hat" on $\tilde{S}(t)$ to simplify the notation, we have [see (2.8-80)]

$$\tilde{S}(t) = \sqrt{\frac{2E_b}{T_b}} \sum_{n=-\infty}^{\infty} \tilde{a}_{0,n} C_0\left(t - nT_b\right) + \sqrt{\frac{2E_b}{T_b}} \sum_{n=-\infty}^{\infty} \tilde{a}_{1,n} C_1\left(t - nT_b\right) \tag{2.8-104}$$

with corresponding received signal

$$\tilde{Z}(t) = \tilde{S}(t) + \tilde{N}(t) \tag{2.8-105}$$

The second term in (2.8-104) may be viewed as an interference term. Since we have restricted ourselves to a linear receiver of the type in Fig. 2-39, we shall

at first ignore this interference term and design the receive filter to match only the first of the two AMP components in (2.8-104). Hence, in view of (2.8-102), we form the output statistics $\mathrm{Re}\{r_{0,2k}\}$ and $\mathrm{Im}\{r_{0,2k+1}\}$, which are obtained by alternately sampling the real and imaginary components of the output of a matched filter (impulse response $h(t) = C_0(t)$) at T_b-s intervals, where $C_0(t)$ is now defined as in (2.8-77). Substituting (2.8-105) together with (2.8-104) into (2.8-103) and simplifying gives

$$\begin{aligned}\mathrm{Re}\{r_{0,2k}\} &= \sqrt{\frac{2E_b}{T_b}}\left[\sum_m \tilde{a}_{0,2k-2m} p_{00}(2mT_b) + \sum_m \tilde{a}_{1,2k-2m+1} p_{10}(2mT_b)\right] \\ &\quad + \mathrm{Re}\{w_{2k}\} \\ \mathrm{Im}\{r_{0,2k+1}\} &= \sqrt{\frac{2E_b}{T_b}}\left[\sum_m \mathrm{Im}\{\tilde{a}_{0,2k-2m+1}\} p_{00}(2mT_b)\right. \\ &\quad \left. + \sum_m \mathrm{Im}\{\tilde{a}_{1,2k-2m+2}\} p_{10}(2mT_b)\right] + \mathrm{Im}\{w_{2k+1}\}\end{aligned} \tag{2.8-106}$$

where

$$\left.\begin{aligned} p_{00}(t) &\triangleq \int C_0(\tau) C_0(\tau - t)\, d\tau \\ p_{10}(t) &\triangleq \int C_1(\tau) C_0(\tau - t)\, d\tau \\ w_k &\triangleq \int \tilde{N}(t) C_0(t - kT)\, dt \end{aligned}\right\} \tag{2.8-107}$$

Notice that even if we ignore the interference term in (2.8-104), i.e., assume the $\tilde{a}_{1,k}$'s are all equal to zero, the metric components in (2.8-106) still contain ISI terms due to the $\tilde{a}_{0,k}$ symbols in that $p_{00}(2mT_b) \neq 0$ for $m \neq 0$. Thus, bit-by-bit decisions based on the $r_{0,k}$'s are not optimum. Furthermore, when the interference term in (2.8-104) is accounted for, then bit-by-bit decisions based on the $r_{0,k}$'s are even more suboptimum. In Ref. 46, it is shown that by applying an MMSE criterion, the performance of the linear receiver can be improved by inserting between the matched filter and the threshold decision device a Weiner

estimator, which takes the form of a finite impulse response (FIR) filter. This should not be too surprising, since it is well-known that such a filter combination is optimum (in the mean-square error sense) for any binary pulse stream that contains ISI and is transmitted over an AWGN. The input-output sample characteristic of the Wiener filter with real coefficients $\{c_k, -N \le k \le N\}$ has the mathematical form

$$y_n = \sum_{k=-N}^{N} c_k r_{0,n-2k} \tag{2.8-108}$$

Equivalently, the transfer function of this filter is given by

$$C\left(e^{j2\pi f(2T_b)}\right) = \sum_{k=-N}^{N} c_k e^{j2\pi fk(2T_b)} \tag{2.8-109}$$

Consequently, bit-by-bit decisions are made using $\mathrm{Re}\{y_{2k}\}$ and $\mathrm{Im}\{y_{2k+1}\}$ in place of $\mathrm{Re}\{r_{0,2k}\}$ and $\mathrm{Im}\{r_{0,2k+1}\}$ in (2.8-102). The evaluation of the coefficients $\{c_k\}$ is performed in Ref. 46 as the solution to a set of Wiener-Hopf (linear) equations involving samples of $p_{00}(t)$ and $p_{10}(t)$, namely,

$$\left.\begin{aligned} \sum_{k=-N}^{N} \Psi_{ik} c_k &= \left(\sqrt{\frac{2E_b}{T_b}}\right)^{-1} p_{00}\left(-2iT_b\right), \quad -N \le i \le N \\ \Psi_{ik} &= \sum_m p_{00}\left(2mT_b\right) p_{00}\big(2\left(m+k-i\right)T_b\big) \\ &\quad + \sum_m p_{10}\big(\left(2m-1\right)T_b\big) p_{10}\big(2\left(m+k-i-1\right)T_b\big) \\ &\quad + \frac{N_0 T_b}{2E_b} p_{00}\left(2\left(k-i\right)T_b\right) \end{aligned}\right\} \tag{2.8-110}$$

Instead of implementing two separate filters, the matched and Wiener filters can be combined into a single optimum filter with impulse response

$$h_o(t) = \mathcal{F}^{-1}\left\{\mathcal{F}\big\{C_0(-t)\big\} C\left(e^{j2\pi f(2T_b)}\right)\right\} = \sum_{k=-\infty}^{\infty} c_k C_0\left(-t+2kT_b\right) \tag{2.8-111}$$

Alternating samples of the real and imaginary parts of the output of $h_o(t)$ in (2.8-111) taken at T_b-s intervals produces the $\mathrm{Re}\{y_{2k}\}$ and $\mathrm{Im}\{y_{2k+1}\}$ values needed for decisions on $\{a_{0,n}\}$. A comparison of the impulse response of the optimum receive filter as given by (2.8-111), with just the matched filter portion, i.e, $h(t) = C_0(-t)$, is illustrated in Fig. 2-40. The eye diagram of the signal at the output of the optimum receive filter is illustrated in Fig. 2-41 for the case of $BT_b = 0.25$.

Upper and lower bounds on the error probability of the linear MSK-type receiver for GMSK are derived in Ref. 46 in the form of the sum of two Gaussian probability integrals with appropriate arguments. These bounds were evaluated for the case of $BT_b = 0.25$ and $N = 11$ FIR filter coefficients. While it is true that the four-state VA receiver performs better than the linear receiver because in the former, the second AMP component [see (2.8-104)] is considered as relevant whereas in the latter, it is treated as interference, the difference in performance between the two is quite small. The reason that the performance difference is small in the GMSK case is because the second AMP component has small energy. For CPM schemes with rational modulation index other than 0.5, one might expect a larger improvement from the simplified VA scheme.

2.8.2.7 Spectral Considerations in the Presence of Data Imbalance. Analogous to the discussion in Sec. 2.7.2, we consider here the spectral behavior of MSK, GMSK, and precoded GMSK in the presence of data imbalance. For linear modulations produced by amplitude modulation of a binary pulse stream on a carrier, e.g., QPSK and OQPSK, the effect of data imbalance on the PSD is well documented, e.g., Chap. 2 of Ref. 1, manifesting itself in the addition of a discrete spectral component to the overall PSD with no effect on the shape of

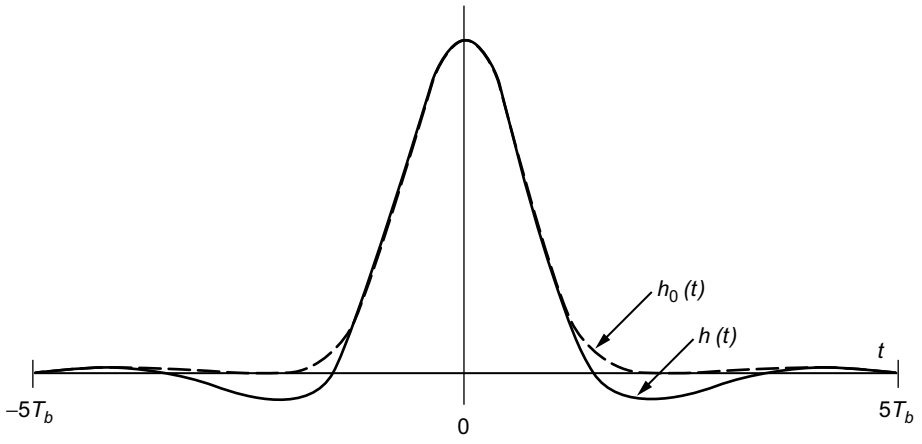

Fig. 2-40. Comparison of the impulse responses of the optimum and matched receive filters. Vertical scaling is normalized. Redrawn from [46].

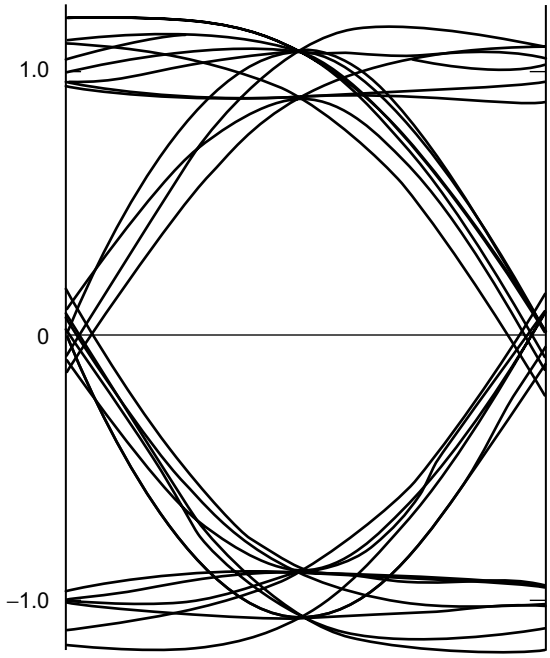

Fig. 2-41. Eye diagram at the output of the optimum receiver filter. Scaling is normalized. Redrawn from [46].

the continuous component. For phase (or frequency) modulation, the evaluation of the PSD is considerably more complex, and the effect of data imbalance is quite different in terms of its impact on both the discrete and continuous spectral components of the modulator output. Because of these important differences and their significance in relation to the specification on the tolerable amount of data imbalance, the presentation will devote more attention to detail, considering first the more generic MSK-type modulations and then including GMSK as a specific case.

Of the many techniques available for evaluating the PSD of CPM schemes [1,15,27,52,59], the one deemed most convenient by the author, particularly for MSK-type modulations with data imbalance, is that which results from the Laurent representation. As previously noted, when the input binary data is random and balanced, the complex data sequences that characterize each of the 2^{L-1} AMP components are themselves uncorrelated, and, furthermore, are uncorrelated with each other. As such, the PSD of the composite CPM waveform is equal to the sum of the PSDs of the AMP components, each of which is computed by conventional PSD evaluation techniques for binary amplitude (unit magnitude)

modulation of a carrier with a complex i.i.d. data sequence. The resulting PSD under these ideal circumstances was presented in Sec. 2.8.2.4.

Here we expand upon the PSD evaluation found in Laurent [52] to include the case of input data imbalance. Specifically, we shall show that because of the presence of data imbalance, the effective complex data sequences that typify each AMP pulse stream are now themselves correlated and, in addition, are correlated with each other. The correlation properties of each of these sequences resemble those of a first-order Markov process, and, hence, the PSD for each contains a factor due to the pulse shape as well as a factor due to the sequence correlation. Likewise, the cross-correlation properties of the sequences contain pulse shape and correlation factors.

We begin by presenting the generic result for the PSD of a modulation composed of a group of correlated data pulse trains, each of which contains its own real pulse shape and complex data stream. Next, we apply this generic PSD formula to first MSK and then GMSK. Since MSK is a full-response scheme, its Laurent representation has only a single pulse stream, and, thus, the PSD has no cross-correlation components. In line with our previous discussions of approximate AMP representations of GMSK, in evaluating the PSD of GMSK, we shall employ the two pulse stream approximation discussed in Sec. 2.8.2.3*a* and characterized by (2.8-80) and later on in (2.8-104). The results that follow are taken primarily from Ref. 60.

a. A Generic Expression for the PSD of a Sum of Random Pulse Trains with Complex Data Symbols. Consider finding the PSD of a complex signal, $\tilde{S}(t)$, of the form in (2.8-67). The traditional method of evaluating such a PSD is to first find the autocorrelation function of $\tilde{S}(t)$, namely, $R_{\tilde{S}}(t, t+\tau) = E\left\{\tilde{S}(t)\,\tilde{S}^*(t+\tau)\right\}$, then time-average to remove the cyclostationary property, and, finally, take the Fourier transform of the result, i.e.,

$$S_{\tilde{S}}(f) = \mathcal{F}\left\{\left\langle R_{\tilde{S}}(t, t+\tau)\right\rangle\right\} \tag{2.8-112}$$

By a straightforward extension of the results in Chap. 2 of Ref. 1, the following result can be obtained:

$$S_{\tilde{S}}(f) = \sum_{i=0}^{2^{L-1}-1} S_{ii}(f) + \sum_{i=0}^{2^{L-1}-1} \sum_{\substack{j=0 \\ i<j}}^{2^{L-1}-1} S_{ji}(f) \tag{2.8-113}$$

where

$$S_{ii}(f) = S_{\tilde{a}_i}(f)\, S_{p_i}(f) \tag{2.8-114}$$

with

$$\left.\begin{aligned} S_{\tilde{a}_i}(f) &= \sum_{l=-\infty}^{\infty} R_{\tilde{a}_i}(l)\, e^{-j2\pi f l T_b}, \quad R_{\tilde{a}_i}(l) = E\left\{\tilde{a}_{i,k}\tilde{a}^*_{i,k+l}\right\} \\ S_{p_i}(f) &= \frac{1}{T_b}\left|P_i(f)\right|^2, \quad P_i(f) \triangleq \mathcal{F}\left\{C_i(t)\right\} \end{aligned}\right\} \tag{2.8-115}$$

and

$$S_{ji}(f) = 2\ \mathrm{Re}\left\{S_{\tilde{a}_{ji}}(f)\, S_{p_{ji}}(f)\right\} \tag{2.8-116}$$

with

$$\left.\begin{aligned} S_{\tilde{a}_{ji}}(f) &= \sum_{l=-\infty}^{\infty} R_{\tilde{a}_{ji}}(l)\, e^{-j2\pi f l T_b}, \quad R_{\tilde{a}_{ji}}(l) = E\left\{\tilde{a}_{j,k}\tilde{a}^*_{i,k+l}\right\} \\ S_{p_{ji}}(f) &= \frac{1}{T_b}P_i(f)\, P_j^*(f), \quad P_i(f) \triangleq \mathcal{F}\left\{C_i(t)\right\} \end{aligned}\right\} \tag{2.8-117}$$

Clearly then, the evaluation of the PSD involves finding the Fourier transform of the various pulse shapes in the AMP representation and both the autocorrelation and cross-correlation of the equivalent complex data sequences.

b. Cross-Correlation Properties of the Equivalent Complex Data Symbols and Evaluation of the PSD. For MSK, the equivalent complex data symbols, $\{\tilde{a}_{0,n}\}$, are defined in terms of the actual input data symbols $\{\alpha_n\}$ by the iterative (complex differential encoding) relation

$$\tilde{a}_{0,n} \triangleq e^{j(\pi/2)A_{0,n}} = j\alpha_n\tilde{a}_{0,n-1} \quad \Rightarrow \quad \tilde{a}_{0,2n} \in \{j,-j\}, \quad \tilde{a}_{0,2n+1} \in \{1,-1\} \tag{2.8-118}$$

Suppose now that $\{\alpha_n\}$ characterizes a random i.i.d. imbalanced source, i.e., one where $\Pr\{\alpha_n = 1\} = 1 - p, \quad \Pr\{\alpha_n = -1\} = p$ with $0 \le p \le 1$. Then, it is straightforward to show that $\{\tilde{a}_{0,n}\}$ is a first-order Markov source and as such, it is balanced, i.e.,

$$\left.\begin{aligned} &\Pr\{\tilde{a}_{0,n}=j\}=\frac{1}{2}, \quad \Pr\{\tilde{a}_{0,n}=-j\}=\frac{1}{2} \quad \text{for } n \text{ even} \\ &\Pr\{\tilde{a}_{0,n}=1\}=\frac{1}{2}, \quad \Pr\{\tilde{a}_{0,n}=-1\}=\frac{1}{2} \quad \text{for } n \text{ odd} \end{aligned}\right\} \tag{2.8-119}$$

and, thus, $E\{\tilde{a}_{0,n}\}=0$. However, while the differential encoding operation converts the imbalanced random i.i.d. source to a balanced source,[22] the symbols of the latter are now correlated. Using the defining relation for $\{\tilde{a}_{0,n}\}$, it is straightforward to show that

$$R_{\tilde{a}_0}(l) \triangleq E\left\{\tilde{a}_{0,n}\tilde{a}^*_{0,n+l}\right\}=[-j(1-2p)]^l, \quad l \text{ integer}, \quad R_{\tilde{a}_0}(-l)=R^*_{\tilde{a}_0}(l) \tag{2.8-120}$$

i.e., $\{\tilde{a}_{0,n}\}$, behaves analogously to a first-order Markov source having a transition probability equal to p. The discrete Fourier transform of (2.8-120) as needed in (2.8-115) is obtained as

$$S_{\tilde{a}_0}(f)=\sum_{l=-\infty}^{\infty} R_{\tilde{a}_0}(l)\, e^{-j2\pi flT_b}=\sum_{l=-\infty}^{\infty}\left[-j(1-2p)\right]^l e^{-j2\pi flT_b}$$

$$=1+2\sum_{l=-\infty}^{\infty}(1-2p)^l\, e^{-j2\pi l(fT_b+[1/4])} \tag{2.8-121}$$

Using a well-known result [61] for the series in (2.8-121), namely,

$$\sum_{k=1}^{\infty} a^k \cos k\theta=\frac{a\cos\theta-a^2}{1-2a\cos\theta+a^2} \tag{2.8-122}$$

we obtain the closed-form result

$$S_{\tilde{a}_0}(f)=\frac{4p(1-p)}{2(1-2p)(1+\sin 2\pi fT_b)+4p^2} \tag{2.8-123}$$

Finally, taking the Fourier transform of the pulse shape $C_0(t)=\sin \pi t/2T_b$ and substituting its squared magnitude in (2.8-115), the complex baseband PSD of MSK with imbalanced data input becomes

[22] The implication of a balanced equivalent complex symbol stream for the AMP representation of MSK is that no discrete spectrum will be generated.

$$S_{\tilde{m}}(f;p) = T_b \frac{16}{\pi^2} \frac{\cos 2\pi f T_b}{\left(1 - 16 f^2 T_b^2\right)^2} \left[\frac{4p(1-p)}{2(1-2p)(1+\sin 2\pi f T_b) + 4p^2} \right] \quad (2.8\text{-}124)$$

Note that because of the presence of the term $\sin 2\pi f T_b$ in the denominator of (2.8-123), the equivalent baseband spectrum of (2.8-124) is not symmetric around $f = 0$. Since the PSD of the true MSK signal is related to the equivalent baseband PSD by

$$S_s(f;p) = \frac{1}{4}\left[S_{\tilde{m}}(f + f_c; p) + S_{\tilde{m}}(-f + f_c; p) \right] \quad (2.8\text{-}125)$$

then, equivalently, the PSD of (2.8-125) will have a tilt around the carrier. Also, since in addition from (2.8-124) we have

$$S_{\tilde{m}}(f; 1-p) = S_{\tilde{m}}(-f; p) \quad (2.8\text{-}126)$$

the tilt of the PSD of (2.8-125) reverses when the probability distribution of the input data is reversed. Finally, for $p = 1/2$, i.e., balanced random data input, the factor in brackets in (2.8-124) becomes equal to unity, and one obtains the well-known two-sided PSD of conventional MSK [see (2.8-35)], which is symmetrical around the origin.

For GMSK, the equivalent complex data symbols, $\{\tilde{a}_{0,n}\}$, are defined in terms of the actual input data symbols $\{\alpha_n\}$ by the iterative relations in the first two equations of (2.8-79). Suppose that $\{\alpha_n\}$ again characterizes a random i.i.d. imbalanced source; then, the autocorrelation function of the first equivalent symbol stream is given by (2.8-120) and its associated discrete Fourier transform by (2.8-123). Thus, the PSD of the first component of the AMP representation of GMSK is

$$S_{00}(f;p) = \frac{1}{T_b} |P_0(f)|^2 \left[\frac{2p(1-p)}{(1-2p)(1+\sin 2\pi f T_b) + 2p^2} \right], \quad P_0(f) \triangleq \mathcal{F}\{C_0(t)\} \quad (2.8\text{-}127)$$

with $C_0(t)$ defined in (2.8-77) and evaluated from (2.8-69) and (2.8-70), using the GMSK phase pulse.

Following a similar procedure as that used to derive (2.8-20), it can be shown that the autocorrelation function of the second equivalent symbol stream (which is also balanced and therefore has zero mean) is given by [60, Appendix]

$$R_{\tilde{a}_1}(l) \triangleq E\left\{\tilde{a}_{1,n}\tilde{a}^*_{1,n+l}\right\} = \begin{cases} 1, & l=0 \\ -j\,(1-2p)^3\,, & l=1, \\ \left[-j\,(1-2p)\right]^l\,, & l\geq 2 \end{cases} \quad R_{\tilde{a}_1}(-l) = R^*_{\tilde{a}_1}(l) \tag{2.8-128}$$

with discrete Fourier transform

$$S_{\tilde{a}_1}(f) = \sum_{l=-\infty}^{\infty} R_{\tilde{a}_1}(l)\, e^{-j2\pi f l T_b} = S_{\tilde{a}_0}(f) + 8p\,(1-2p)\,(1-p)\sin 2\pi f T_b \tag{2.8-129}$$

Therefore, the PSD of the second component of the AMP representation of GMSK is

$$S_{11}(f;p) = \frac{1}{T_b}\left|P_1(f)\right|^2 4p\,(1-p) \times \left[\frac{1}{2\,(1-2p)\,(1+\sin 2\pi f T_b)+4p^2} + 2\,(1-2p)\sin 2\pi f T_b\right], \tag{2.8-130}$$

$$P_1(f) \triangleq \mathcal{F}\left\{C_1(t)\right\}$$

Note again that because of the presence of the term $\sin 2\pi f T_b$ in the denominator of (2.8-130), the equivalent baseband spectrum is not symmetric around $f = 0$.

What remains is to compute the cross-correlation function of the two equivalent complex symbol streams. Following the same procedure as for obtaining the autocorrelation function of the individual pulse streams, we obtain [60, Appendix]

$$R_{\tilde{a}_{10}}(l) \triangleq E\left\{\tilde{a}_{1,n}\tilde{a}^*_{0,n+l}\right\} = \begin{cases} \left[-j(1-2p)\right]^{l+1}\,, & l\geq 0 \\ (1-2p)^2, & l=-1\,, \\ \left[j(1-2p)\right]^{-(l+1)}\,, & l\leq -2 \end{cases} \quad R_{\tilde{a}_{01}}(-l) = R^*_{\tilde{a}_{10}}(l) \tag{2.8-131}$$

with discrete Fourier transform

$$S_{\tilde{a}_{10}}(f) = \sum_{l=-\infty}^{\infty} R_{\tilde{a}_{10}}(l)\, e^{-j2\pi f l T_b} = e^{j2\pi f T_b}\left[S_{\tilde{a}_0}(f) - 4p(1-p)\right] \quad (2.8\text{-}132)$$

Thus, the cross-spectrum of $\tilde{m}(t)$ is from (2.8-116)

$$S_{10}(f;p) =$$

$$8p(1-p)\,\mathrm{Re}\left\{\left[\frac{1}{2(1-2p)(1+\sin 2\pi f T_b)+4p^2} - 1\right] e^{j2\pi f T_b}\frac{1}{T_b}P_0(f)\,P_1^*(f)\right\} \quad (2.8\text{-}133)$$

which is also not symmetric around $f = 0$. Finally, the complex baseband PSD of GMSK (based on the two pulse stream AMP approximation) with imbalanced data input becomes

$$S_{\tilde{m}}(f;p) = S_{00}(f;p) + S_{11}(f;p) + S_{10}(f;p) \quad (2.8\text{-}134)$$

where $S_{00}(f;p), S_{11}(f;p)$, and $S_{10}(f;p)$ are defined in (2.8-127), (2.8-130), and (2.8-133), respectively.

Before proceeding, we point out that with some additional computation (which would be warranted if one were interested in very low PSD levels), the PSD evaluation procedure discussed above can be extended to include more than just the first two (dominant) AMP pulse streams. In fact, the results of Sec. 2.8.2.7*a* are quite general and, analogous to (2.8-120), (2.8-128), and (2.8-131), all one needs to compute are the autocorrelation and cross-correlation functions of the remainder of the equivalent data symbol streams, e.g., see (2.8-79) for $L = 4$.

c. Precoded MSK and GMSK. As previously discussed in Secs. 2.8.1.3 and 2.8.2.3*b*, conventional I-Q-type receivers for MSK and GMSK modulations suffer a small performance penalty due to the differential encoding operation inherent in these modulations and, thus, the need for differential decoding at the receiver. Precoding the input data with a differential decoder removes the need for differential decoding at the receiver and, thus, eliminates this penalty. From a spectral standpoint, this precoding operation has no effect on the PSD of the transmitted signal when the input data are balanced. However, when the input data are imbalanced, as is the case of interest in this section, the precoder has a definite effect on the transmitted signal PSD. To see how this comes about, we shall first consider the simpler case of MSK.

Suppose that prior to phase modulation of the carrier the input data stream $\boldsymbol{\alpha} = (\cdots, \alpha_{-2}, \alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \cdots)$ is first converted to a complex data stream via

$$\alpha'_n = \alpha_n j^n \tag{2.8-135}$$

and then passed through a differential decoder that satisfies the recursion relation

$$\beta_n = -j\alpha'_n \left(\alpha'_{n-1}\right)^* \tag{2.8-136}$$

where β_n denotes the complex binary output of the decoder (input to the MSK modulator) in the nth bit interval. Substituting (2.8-135) into (2.8-136), we see that

$$\beta_n = -j\left(\alpha_n j^n\right)\left(\alpha_{n-1}\left(-j\right)^{n-1}\right) = \alpha_n \alpha_{n-1} \tag{2.8-137}$$

which is a conventional differential decoding of the true input data bits. Since the cascade of the MSK differential encoding relationship $\tilde{a}_{0,n} = j\alpha_n \tilde{a}_{0,n-1}$ and the differential decoder of (2.8-136) produces a unity gain transmission path, i.e.,

$$\beta_n = -j\tilde{a}_{0,n}\tilde{a}^*_{0,n-1} = -j\left(j\alpha_n \tilde{a}_{0,n-1}\right)\tilde{a}^*_{0,n-1} = \alpha_n \left|\tilde{a}_{0,n-1}\right|^2 = \alpha_n \tag{2.8-138}$$

one can deduce that for an input binary complex i.i.d. bit sequence, $\boldsymbol{\alpha}' = \left(\cdots, \alpha'_{-2}, \alpha'_{-1}, \alpha'_0, \alpha'_1, \alpha'_2, \cdots\right)$, as in (2.8-135), precoded MSK using the precoder (differential decoder) in (2.8-136) is exactly the same as a Laurent representation of MSK (a single, complex symbol pulse stream with half-sinusoidal pulse shape) with the same input data sequence, i.e., $\{\tilde{a}_{0,n}\} = \boldsymbol{\alpha}'$.

The consequence of the above equivalence is that since the conversion from $\boldsymbol{\alpha}$ to $\boldsymbol{\alpha}'$ does not change the statistical (correlation) properties of the sequence, then based on the Laurent AMP representation, we conclude that the PSD of precoded MSK is that of a linear modulation with an i.i.d. uncorrelated complex imbalanced data source and as such, has a continuous component given by

$$S_{\text{P-MSK}}(f)\,|_{cont.} = 4p(1-p)T_b \frac{16}{\pi^2} \frac{\cos^2 2\pi f T_b}{\left(1 - 16f^2 T_b^2\right)^2} \tag{2.8-139}$$

and a discrete component derived analogously to the results in Chap. 2 of Ref. 1 as

$$S_{\text{P-MSK}}(f)|_{discr.} = (1-2p)^2 \sum_{k=-\infty}^{\infty} \frac{4}{\pi^2} \frac{1}{(1-4k^2)^2} \delta\left(f - \frac{k}{2T_b}\right) \qquad (2.8\text{-}140)$$

where P-MSK denotes precoded MSK. Thus, in summary, the addition of a precoder to the input of an MSK modulator with imbalanced data input removes the tilt of the MSK spectrum due to the imbalance and replaces it with a discrete spectral component, as is typical of linear modulations.

For precoded GMSK approximated by the first two AMP components, the PSD in the presence of data imbalance was derived in Ref. 62. Without going into great detail, the resulting expressions for the continuous and discrete PSD components are given below:

$$\begin{aligned} &S_{\text{P-GMSK}}(f)|_{cont.} = \\ &4p(1-p)\frac{1}{T_b}|P_0(f)|^2 + \left\{1-(1-2p)^6 + 2\left[(1-2p)^2 - (1-2p)^6\right]\cos 2\pi f T_b \right. \\ &\left. + 2\left[(1-2p)^4 - (1-2p)^6\right]\cos 4\pi f T_b\right\}\frac{1}{T_b}|P_1(f)|^2 \\ &+ 2\,\text{Re}\left\{\left[(1-2p)^2 - (1-2p)^4\right]\left[1+\exp(-2\pi f T_b) + \exp(-4\pi f T_b)\right]\right\} \\ &\times \frac{1}{T_b} P_0(f) P_1^*(f) \end{aligned} \qquad (2.8\text{-}141a)$$

$$\begin{aligned} &S_{\text{P-GMSK}}(f)|_{discr.} = \\ &\left[(1-2p)^2 \sum_{k=-\infty}^{\infty}\left(\frac{1}{2T_b}\right)^2 \left|P_0\left(\frac{k}{2T_b}\right)\right|^2 + (1-2p)^6 \sum_{k=-\infty}^{\infty}\left(\frac{1}{2T_b}\right)^2 \left|P_1\left(\frac{k}{2T_b}\right)\right|^2 \right. \\ &\left. +2(1-2p)^2 \sum_{k=-\infty}^{\infty}\left(\frac{1}{2T_b}\right)^2 \text{Re}\left\{P_0\left(\frac{k}{2T_b}\right)P_1^*\left(\frac{k}{2T_b}\right)\right\}\right]\delta\left(f-\frac{k}{2T_b}\right) \end{aligned} \qquad (2.8\text{-}141b)$$

where, as before, $P_0(f)$ and $P_1(f)$ are the Fourier transforms of the AMP pulse shapes $C_0(t)$ and $C_1(t)$. As was true for precoded MSK, applying a precoder to

a GMSK modulator with imbalanced data input removes the tilt of the GMSK spectrum due to the imbalance and replaces it with a discrete spectral component. Figures 2-42 and 2-43 illustrate the PSD of precoded GMSK as computed from the sum of (2.8-141a) and (2.8-141b) for $BT_b = 0.25$ with 10 percent data imbalance and 60 percent data imbalance ($p = 0.2$), respectively. Included with the theoretical results are numerical results obtained from computer simulation [62]. As can be seen from these illustrations, the theory matches very closely the simulation results.

2.8.2.8 Synchronization Techniques. In addition to the previously discussed advantages of the AMP representation in so far as spectrum evaluation and ideal receiver implementation, there is yet another advantage having to do with carrier synchronization of the receiver. Mengali and Andrea [63] discuss the use of the Laurent representation for CPM primarily in the context of the single pulse stream approximation and, as such, arrive at decision-directed phase estimation structures that are analogous to those used for MSK. Similar decision-directed (data-aided) methods of obtaining symbol time and carrier phase tracking estimates for precoded CPM (in particular, GMSK) were also considered in Ref. 55.

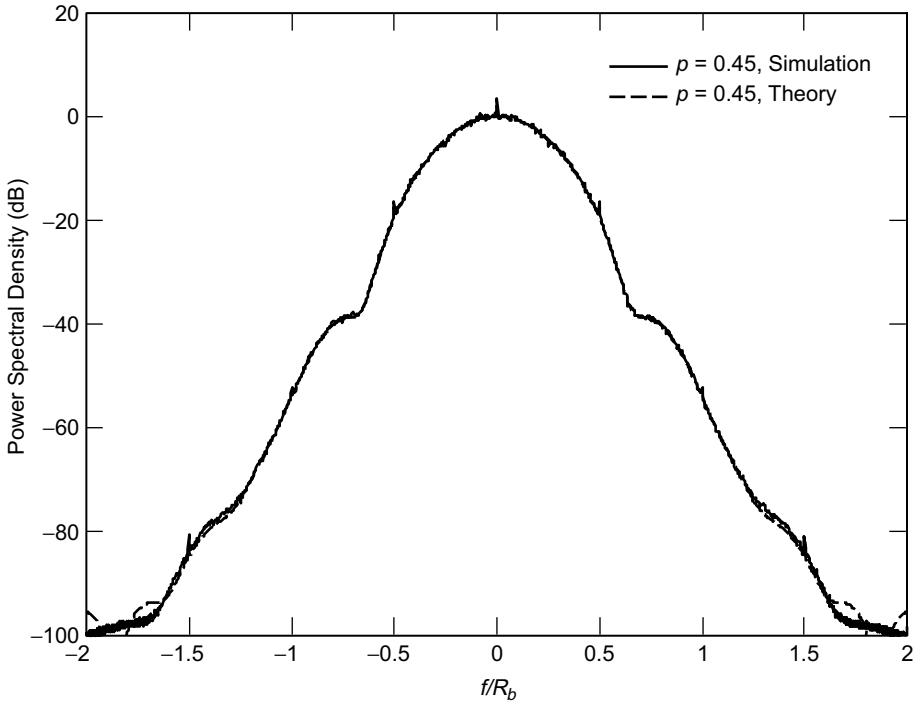

Fig. 2-42. GMSK spectrum with precoding and 10 percent data imbalance.

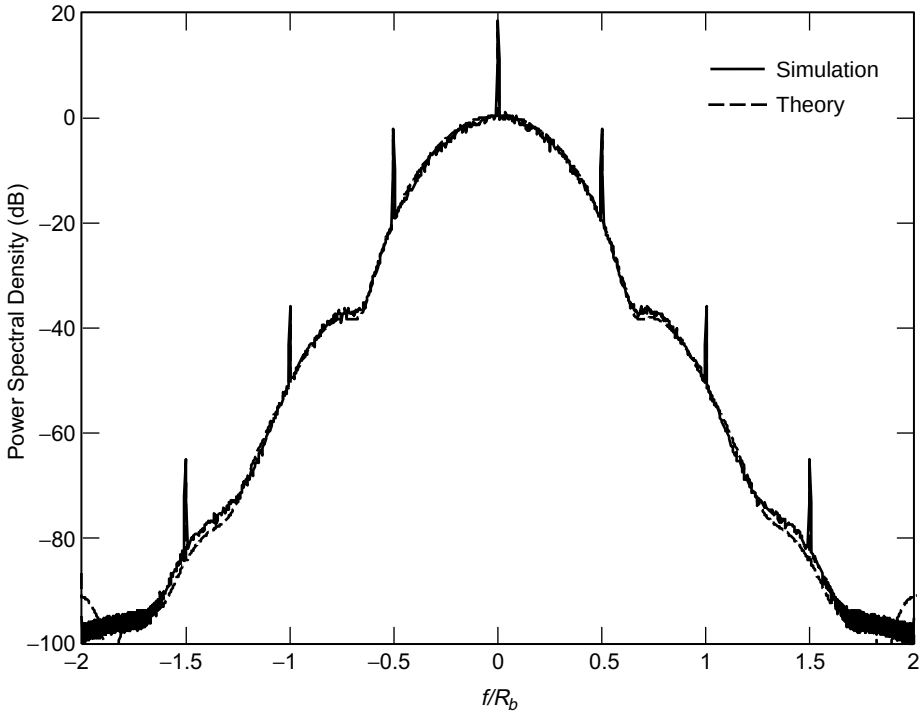

Fig. 2-43. GMSK spectrum with precoding and 60 percent data imbalance.

In this section of the monograph, we extend the carrier synchronization problem two steps further, with the goal of achieving a better solution. First, we consider the two pulse stream approximation suggested by Kaleh [46] rather than the single (main) pulse approximation. Second, using the MAP approach for carrier phase estimation as applied to pulse stream modulations with ISI [64,65], we arrive at an optimum[23] closed-loop structure that is not limited to a decision-directed form and, moreover, accounts for the ISI directly within its implementation. Finally, the tracking performance of this optimum structure is evaluated in terms of its mean-square phase error. Some of the results to be presented here are extracted from Ref. 66.

a. MAP Estimation of Carrier Phase. Consider the received signal, $y(t)$, composed of the sum of $s(t;\theta)$ and AWGN, $n(t)$ (with single-sided PSD, N_0 watts/hertz), where $s(t;\theta)$ is given by (2.8-85) with the addition of a

[23] By optimum we mean that closed-loop structure whose error signal is motivated by the derivative of the log-likelihood ratio associated with the MAP estimation of carrier phase.

uniformly distributed phase, θ, included in each carrier component. Based on an observation of $y(t)$ over the interval $0 \le t \le T_0$, where we arbitrarily assume that T_0 is an even integer multiple, say K_b, of the bit time, T_b ($K_s = K_b/2$ is then an integer multiple of the symbol time, $T_s = 2T_b$), we wish to find the MAP estimate of θ, i.e., the value of θ that maximizes the a posteriori probability, $p(\theta\,|y(t)\,)$, or since θ is assumed to be uniformly distributed, the value of θ that maximizes the conditional probability $p(y(t)\,|\theta\,)$. For an AWGN channel with single-sided noise power spectral density N_0 watts/hertz, $p(y(t)\,|\theta\,)$ has the form

$$p\left(y(t)\,\middle|\,\mathbf{a}_0^I, \mathbf{a}_0^Q, \mathbf{a}_1^I, \mathbf{a}_1^Q, \theta\right) = C\exp\left(-\frac{1}{N_0}\int_0^{T_0}\big(y(t) - s\,(t;\theta)\big)^2 dt\right) \quad \text{(2.8-142)}$$

where C is a normalization constant, and we have added to the conditioning notation the implicit dependence of $s(t;\theta)$ on the i.i.d. I and Q data sequences of the two pulse streams. For a constant envelope (energy) signal such as GMSK, it is sufficient to consider only the term involving the correlation of $y(t)$ and $s(t;\theta)$ and lump the remaining terms into the normalization constant.[24] Thus, we rewrite (2.8-142) as

$$p\left(y(t)\,\middle|\,\mathbf{a}_0^I, \mathbf{a}_0^Q, \mathbf{a}_1^I, \mathbf{a}_1^Q, \theta\right) = C\exp\left(\frac{2}{N_0}\int_0^{T_0} y(t)s\,(t;\theta)\,dt\right) \quad \text{(2.8-143)}$$

where for convenience, we still use C to denote the normalization constant.

Evaluation of (2.8-143) for $s(t;\theta)$ corresponding to a single binary pulse stream, e.g., BPSK, with ISI was considered in Refs. 64 and 65. Extension of the result to an $s(t;\theta)$ corresponding to a single pair of quadrature binary pulse streams (such as QPSK) with identical ISI on the I and Q channels is straightforward and was somewhat discussed in Ref. 64. What we have for the AMP representation of GMSK in (2.8-85) is two pairs of offset quadrature binary pulse streams, each pair having different amounts of ISI. (Recall that $C_0(t)$ is a pulse of width $5T_b$, and $C_1(t)$ is a completely different pulse of width $3T_b$.)

---

[24] We note that for the general ISI problem as treated in Refs. 64 and 65, the energy-dependent exponential term $\exp\{-(1/N_0)\int_0^{T_0} s^2(t;\theta)dt\}$ is not constant and, in fact, depends on the data sequence. However, for the "exact" representation of GMSK by the two pulse stream AMP form, we can make the constant envelope assumption and hence ignore the energy-dependent term.

Without belaboring the details, following substitution of (2.8-85) into (2.8-143) and averaging over the four i.i.d. component data sequences $\mathbf{a}_0^I, \mathbf{a}_0^Q, \mathbf{a}_1^I, \mathbf{a}_1^Q$, we obtain [66]

$$p\left(y(t)\,|\theta\right) = C \prod_{\substack{k=-3 \\ k \text{ odd}}}^{K_b-1} \cosh\left\{I_c\left(k,0,\theta\right)\right\} \prod_{\substack{k=-4 \\ k \text{ even}}}^{K_b-2} \cosh\left\{I_s\left(k,0,\theta\right)\right\}$$

$$\times \prod_{\substack{k=-2 \\ k \text{ even}}}^{K_b-2} \cosh\left\{I_c\left(k,1,\theta\right)\right\} \prod_{\substack{k=-1 \\ k \text{ odd}}}^{K_b-1} \cosh\left\{I_s\left(k,1,\theta\right)\right\} \tag{2.8-144}$$

where

$$\left.\begin{aligned} I_c\left(k,l,\theta\right) &\triangleq \frac{2\sqrt{2E_b/T_b}}{N_0} \int_0^{K_bT_b} r(t)\cos\left(\omega_c t+\theta\right) C_l\left(t-kT_b\right) dt \\ I_s\left(k,l,\theta\right) &\triangleq \frac{2\sqrt{2E_b/T_b}}{N_0} \int_0^{K_bT_b} r(t)\sin\left(\omega_c t+\theta\right) C_l\left(t-kT_b\right) dt \end{aligned}\right\} \tag{2.8-145}$$

Note that because of the presence of ISI in each of the component pulse streams, the arguments of the hyperbolic cosine terms involve integration over the entire observation interval $0 \le t \le K_bT_b$ rather than just integration over a single bit interval, as is customary in such problems when ISI is absent. (Actually the finite duration of $C_0\left(t-kT_b\right)$ and $C_1\left(t-kT_b\right)$ will truncate these integrations to an interval (depending on the value of k) smaller than the observation time interval but still larger than the bit interval.) Finally, the MAP estimate of θ, i.e., θ_{MAP}, is the value of θ that maximizes (2.8-144).

b. Closed-Loop Carrier Synchronization of GMSK. As has been done many times in the past to arrive at closed-loop carrier synchronizers based on open-loop MAP estimates, one takes the natural logarithm of the likelihood ratio, differentiates it with respect to θ, and then uses this as the error signal, $e\left(\theta\right)$, in a closed-loop configuration. The reasoning behind this approach is that $e\left(\theta\right)$ will be equal to zero when $\theta = \theta_{\text{MAP}}$ and, thus, the closed loop will null at the point corresponding to the open MAP phase estimate. Proceeding in this fashion, we obtain

$$e(\theta) \triangleq \frac{d}{d\theta} \ln p\left(y(t) \mid \theta\right)$$

$$= \sum_{\substack{k=-3 \\ k \text{ odd}}}^{K_b-1} I_s(k,0,\theta) \tanh\left\{I_c(k,0,\theta)\right\} - \sum_{\substack{k=-4 \\ k \text{ even}}}^{K_b-2} I_c(k,0,\theta) \tanh\left\{I_s(k,0,\theta)\right\}$$

$$+ \sum_{\substack{k=-2 \\ k \text{ even}}}^{K_b-2} I_s(k,1,\theta) \tanh\left\{I_c(k,1,\theta)\right\} - \sum_{\substack{k=-1 \\ k \text{ odd}}}^{K_b-1} I_c(k,1,\theta) \tanh\left\{I_s(k,1,\theta)\right\}$$

$$\triangleq e_0(\theta) + e_1(\theta) \tag{2.8-146}$$

where we have made use of the fact that from (2.8-145) $I_c(k,l,\theta)$ and $I_s(k,l,\theta)$ are derivatives of each other.

The result in (2.8-146) suggests a superposition of two loops, each contributing a component to the error signal corresponding to associated pulse stream in the two pulse stream AMP representation of GMSK. Figures 2-44(a) and 2-44(b) illustrate the two loop components that must be superimposed to arrive at the closed-loop GMSK carrier synchronizer suggested by the error signal in (2.8-146).[25] We offer this scheme as the "optimum" (in the sense of being MAP motivated) GMSK carrier synchronizer. As is customary, the tanh nonlinearity can be approximated by a linear or hard limiter device for low and high SNR applications, respectively. The rate at which the loop updates its carrier phase estimate can vary from every T_b to every K_bT_b seconds. In the case of the latter extreme, the observation intervals used for each carrier phase estimate do not overlap and, as such, the loop represents a sequential block-by-block implementation of the MAP open-loop estimator. In the case of the former extreme, the observation intervals used for each carrier phase estimate overlap by $(K_b-1)T_b$ s and, as such, the loop represents a sliding window MAP phase estimator.

c. Performance of the GMSK Loop Based on a Single Pulse AMP Representation. In this section, we consider the mean-square error performance of the previously derived closed loop, using just a single pulse stream for the AMP representation of GMSK. As such, the error signal is described by only the first two out of the four terms in (2.8-146), which leads to the implementation in Fig. 2-44(a), i.e., there is no contribution to the error signal from Fig. 2-44(b).

[25] A value of $K_b = 6$ (for any larger value, the noise-free (signal) components of $I_c(k,l,\theta)$ and $I_s(k,l,\theta)$ would not change due to the truncation of the integral caused by the time limitation of $C_0(t-kT_b)$, and $C_1(t-kT_b)$) is no doubt sufficient for these figures.

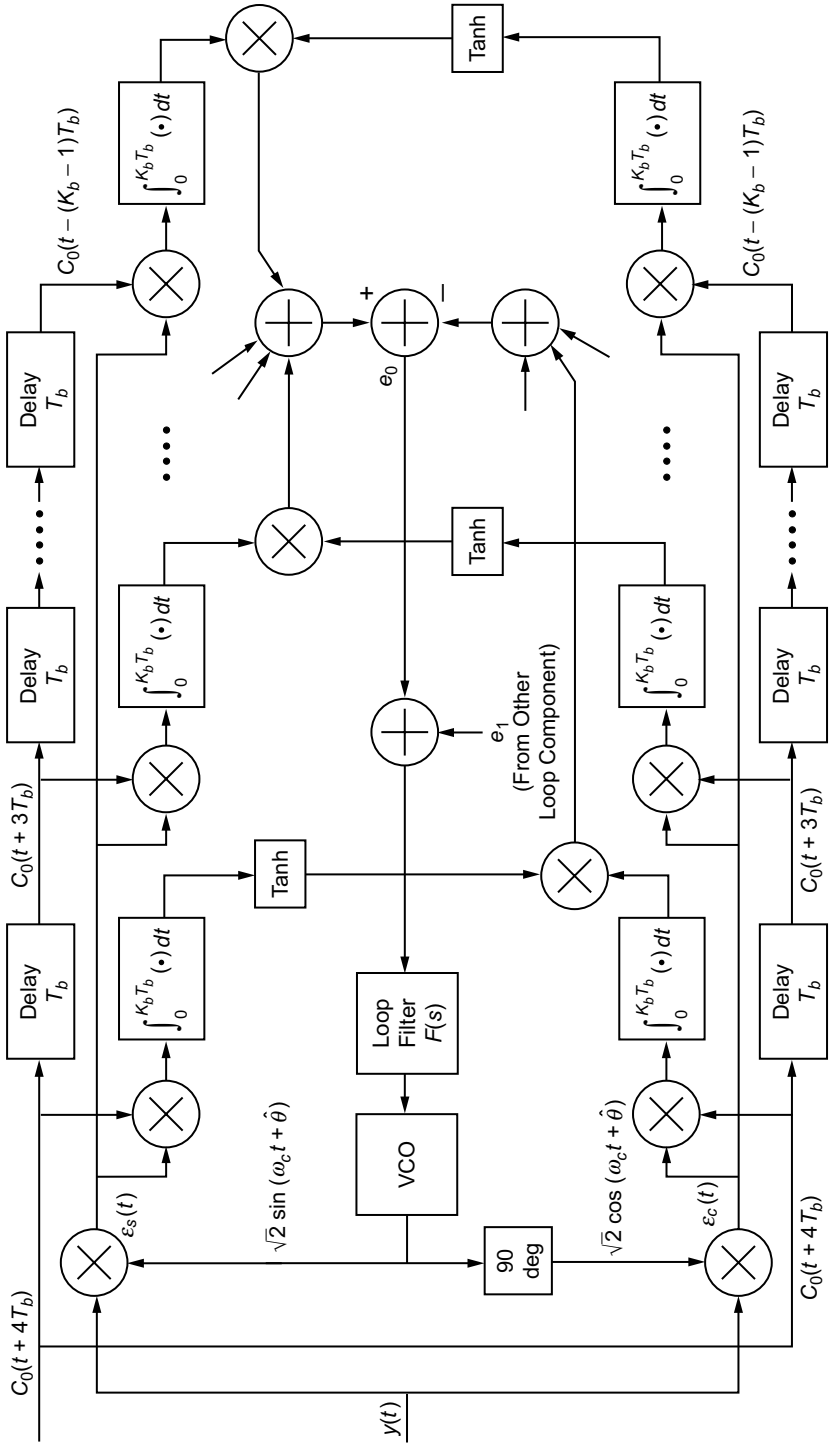

Fig. 2-44(a). Block diagram of a suboptimum ISI-compensated MAP estimation loop for GMSK (first signal component).

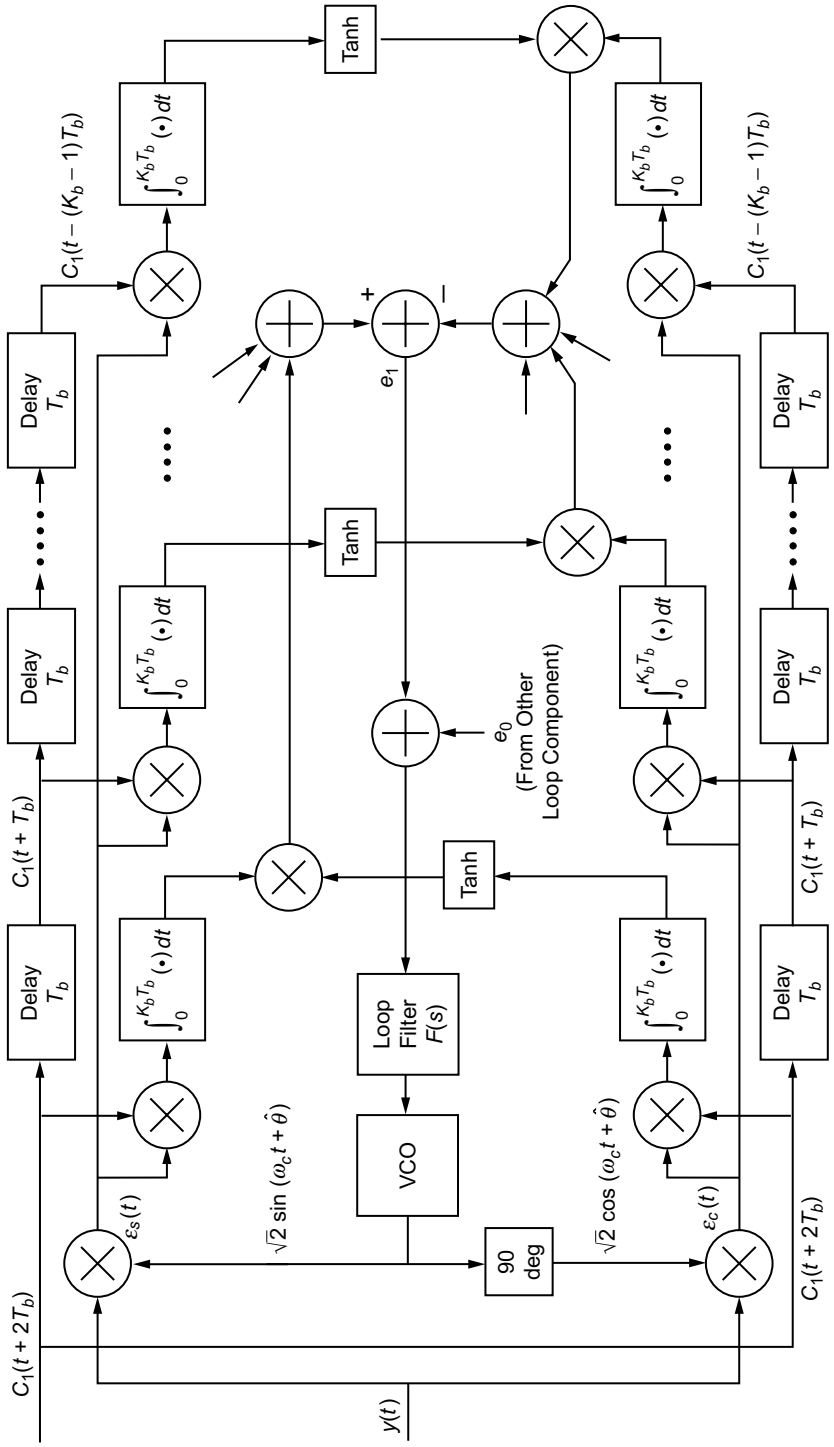

Fig. 2-44(b). Block diagram of a suboptimum ISI-compensated MAP estimation loop for GMSK (second signal component).

In evaluating the performance, we shall consider the linear loop model, wherein the tanh nonlinearity is replaced by a linear function.

To obtain the mean-square phase error performance, we follow the approach taken in [11,67], resulting in the expression

$$\sigma_{2\phi}^2 = \frac{N_E B_L}{K_g^2} \tag{2.8-147}$$

where B_L denotes the loop bandwidth, N_E is the flat single-sided PSD of the equivalent noise process perturbing the loop, and K_g is the slope (with respect to 2ϕ) of the loop S-curve at the origin. Without going into great detail, K_g is obtained as

$$K_g = PT_s^2 \sum_{i=-K_s+1}^{2} \sum_{j=-K_s+1}^{2} \left[-I_{i-(1/2),j}^2 + I_{i-(1/2),j-(1/2)}^2 + I_{i,j}^2 - I_{i,j-(1/2)}^2 \right] \tag{2.8-148}$$

where $P = E_b/T_b$ is the signal power, and the $I_{i,j}$'s are ISI parameters defined by

$$I_{i,j} \triangleq \frac{1}{T_s} \int_0^{K_s T_s} C_0(t + iT_s) C_0(t + jT_s) dt = I_{j,i} \tag{2.8-149}$$

Furthermore, N_E is evaluated as

$$N_E = 2N_0^2 T_s^3 \beta + 4PN_0 T_s^4 \alpha \tag{2.8-150}$$

where $T_s = 2T_b$ is again the effective symbol rate in each of the quadrature channels and the coefficients α and β are given as follows:

$$\begin{aligned} \alpha = & \sum_{i=-K_s+1}^{2} \sum_{l=-K_s+1}^{2} \left[I_{i-(1/2),l-(1/2)}^2 + I_{i,l}^2 - 2I_{i,l-(1/2)}^2 \right] \\ & + 2 \sum_{n=1}^{K_s-1} \sum_{i=-K_s+1}^{2} \sum_{l=-K_s+1}^{2} \\ & \times \left[J_{i-(1/2),l-(1/2)}^2(n) + J_{i,l}^2(n) - J_{i,l-(1/2)}^2(n) - J_{i-(1/2),l}^2(n) \right] \end{aligned} \tag{2.8-151}$$

and

$$\beta = \sum_{i=-K_s+1}^{2} \sum_{l=-(K_s+n)+1}^{2-n} \Bigg[I_{i-(1/2),l-(1/2)}$$

$$\times \sum_{j=-K_s+1}^{2} \left(I_{i-(1/2),j-(1/2)} I_{l-(1/2),j-(1/2)} + I_{i-(1/2),j} I_{l-(1/2),j} \right)$$

$$- I_{i-(1/2),l} \sum_{j=-K_s+1}^{2} \left(I_{i-(1/2),j-(1/2)} I_{l,j-(1/2)} + I_{i-(1/2),j} I_{l,j} \right)$$

$$+ I_{i,l} \sum_{j=-K_s+1}^{2} \left(I_{i,j} I_{l,j} + I_{i,j-(1/2)} I_{l,j-(1/2)} \right)$$

$$- I_{i,l-(1/2)} \sum_{j=-K_s+1}^{2} \left(I_{i,j} I_{l-(1/2),j} + I_{i,j-(1/2)} I_{l-(1/2),j-(1/2)} \right) \Bigg]$$

$$+ 2 \sum_{n=1}^{K_s-1} \sum_{i=-K_s+1}^{2} \sum_{l=-(K_s+n)+1}^{2-n} \Bigg[J_{i-(1/2),l-(1/2)}(n)$$

$$\times \sum_{j=-K_s+1}^{2} \left(I_{i-(1/2),j-(1/2)} I_{l-(1/2),j-(1/2)}(n) + I_{i-(1/2),j} I_{l-(1/2),j}(n) \right)$$

$$- J_{i-(1/2),l}(n) \sum_{j=-K_s+1}^{2} \left(I_{i-(1/2),j-(1/2)} I_{l,j-(1/2)}(n) + I_{i-(1/2),j} I_{l,j}(n) \right)$$

$$+ J_{i,l}(n) \sum_{j=-K_s+1}^{2} \left(I_{i,j} I_{l,j}(n) + I_{i,j-(1/2)} I_{l,j-(1/2)}(n) \right)$$

$$- J_{i,l-(1/2)}(n) \sum_{j=-K_s+1}^{2} \left(I_{i,j} I_{l-(1/2),j}(n) + I_{i,j-(1/2)} I_{l-(1/2),j-(1/2)}(n) \right) \Bigg] \tag{2.8-152}$$

where the additional ISI parameters are defined by

$$I_{i,j}(k) \triangleq \frac{1}{T_s}\int_{kT_s}^{(k+K_s)T_s} C_0(t+iT_s)C_0(t+jT_s)dt = I_{j,i}(k) \tag{2.8-153}$$

and

$$J_{i,j}(k) \triangleq \frac{1}{T_s}\int_{kT_s}^{K_sT_s} C_0(t+iT_s)C_0(t+jT_s)dt = J_{j,i}(k) \tag{2.8-154}$$

which differs from $I_{i,j}(k)$ of (2.8-153) only in that the upper limit is kept fixed at $K_sT_s = K_bT_b$, independent of k. Note also that $J_{i,j}(0) = I_{i,j}(0) = I_{i,j}$, as defined in (2.8-149).

It is customary to rewrite (2.8-147) in the form[26]

$$\sigma_{2\phi}^2 = \frac{4N_0B_L}{PS_L} = \frac{4}{\rho_{\text{PLL}}\mathcal{S}_L} \tag{2.8-155}$$

where $\rho_{\text{PLL}} = P/N_0B_L$ denotes the loop SNR for a phase-locked loop (PLL) and $\mathcal{S}_L$ denotes the so-called squaring loss, which represents the additional degradation in loop SNR caused by the presence of $S\times S$, $S\times N$, and $N\times N$ components in the error signal. Combining (2.8-147) and (2.8-153), we obtain the following expression for the squaring loss:

$$\mathcal{S}_L = \frac{\left\{\sum_{i=-K_s+1}^{2}\ \sum_{j=-K_s+1}^{2}\left[-I_{i-(1/2),j}^2 + I_{i-(1/2),j-(1/2)}^2 + I_{i,j}^2 - I_{i,j-(1/2)}^2\right]\right\}^2}{\dfrac{N_E}{4PN_0T_s^4}} \tag{2.8-156}$$

From (2.8-150), it is possible to write the equivalent normalized flat noise spectral density as

$$\frac{N_E}{4PN_0T_s^4} = \alpha + \frac{\beta}{2PT_s/N_0} = \alpha + \frac{\beta}{2E_s/N_0} \tag{2.8-157}$$

[26] The factor of "4" in (2.8-155) comes from the fact that we are characterizing the variance of the 2ϕ process rather than the ϕ process.

where E_s/N_0 is the symbol energy-to-noise ratio. Finally, using (2.8-157) in (2.8-156), we get the desired result for the squaring loss, namely,

$$\mathcal{S}_L = \frac{\left\{ \sum_{i=-K_s+1}^{2} \sum_{j=-K_s+1}^{2} \left[-I^2_{i-(1/2),j} + I^2_{i-(1/2),j-(1/2)} + I^2_{i,j} - I^2_{i,j-(1/2)} \right] \right\}^2}{\alpha + \dfrac{\beta}{2E_s/N_0}} \tag{2.8-158}$$

which is expressed entirely in terms of the symbol energy-to-noise ratio and the ISI parameters defined in (2.8-149), (2.8-153), and (2.8-154), all of which can easily be computed from knowledge of the main pulse shape, $C_0(t)$.

For GMSK with $BT_b = 0.25$ (equivalently $L = 4$) corresponding to the pulse shape $C_0(t)$ shown in Fig. 2-32, Fig. 2-45 illustrates the squaring loss (in dB) as computed from (2.8-158) versus $E_b/N_0 = (1/2)E_s/N_0$ (in dB) with K_s, the number of symbols in the observation interval, as a parameter. Because the dominant pulse in the AMP representation of GMSK is 5 bits (2.5 symbols) long, it appears that extending the observation beyond the duration of the pulse (i.e., values of $K_s > 3$) provides no further improvement in performance. In fact, the results for $K_s = 2$ and $K_s = 3$ are virtually indistinguishable from

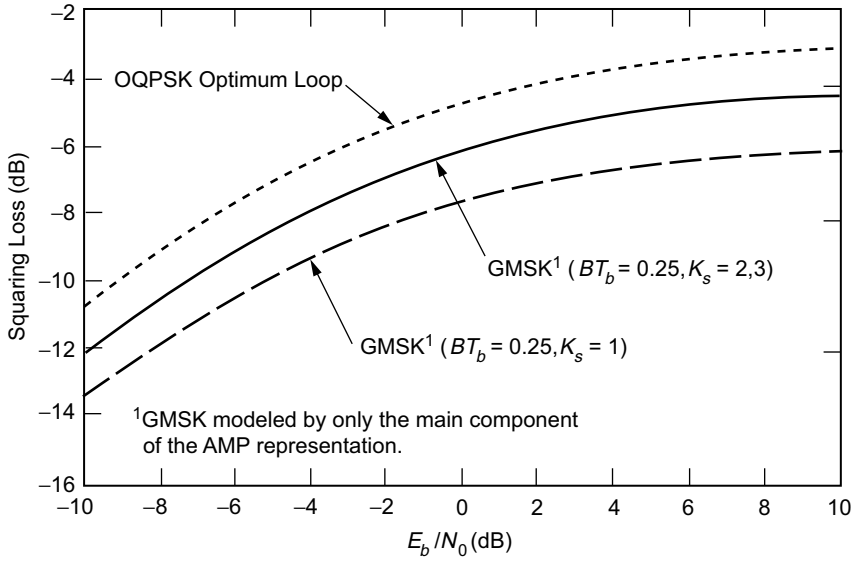

Fig. 2-45. The squaring loss performance of OQPSK and GMSK loops.

each other. Thus, for the chosen value of BT_b, implementing the loop based on a value of $K_s = 2$ is sufficient, thereby reducing the implementation complexity, which increases with the value of K_s. Note, however, that there is a significant improvement in performance by building the structure with $K_s = 2\,(K_b = 4)$ rather than $K_s = 1\,(K_b = 2)$.

Also shown in Fig. 2-45 for purpose of comparison is the performance of the corresponding MAP-motivated (optimum) OQPSK carrier synchronization loop, as obtained from the results in Ref. 12, which employs square pulses of duration T_s and so does not suffer from ISI. Although the OQPSK loop outperforms that of GMSK, we see that the difference between the two (in terms of squaring loss or, equivalently, in terms of equivalent loop SNR) is only a little more than 1 dB. This difference appears to be constant across a large range of E_b/N_0 values (-10 dB to 10 dB) and is a small price to pay for the large improvement in bandwidth efficiency that GMSK affords over OQPSK. Of particular importance is that the loop will, in fact, acquire and track a GMSK modulation at very low E_b/N_0 values, which is important in applications where high-power, efficient, error-correction coding, e.g., convolutional or turbo coding, is added to the system.

Since squaring loss is not a physical quantity that can be determined from computer simulation, to demonstrate the excellent agreement between simulation and analysis, Fig. 2-46 directly plots the equivalent linear loop SNR [i.e.,

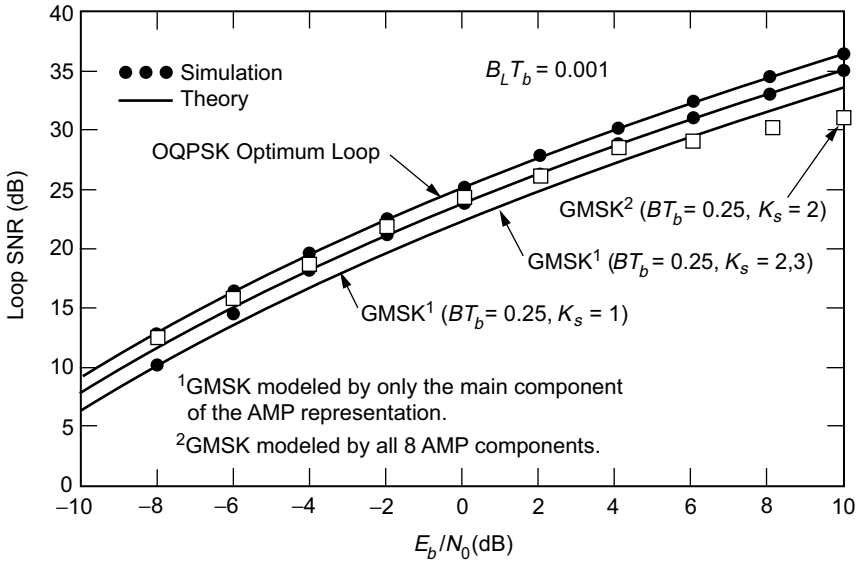

Fig. 2-46. Loop SNR performance of OQPSK and GMSK loops.

the reciprocal of the mean-square phase error as computed from (2.8-155)] versus E_b/N_0 (in dB) for the same parameter values as in Fig. 2-45 and a loop bandwidth-bit time product, $B_L T_b = 0.001$. Here, several different GMSK options were investigated. All of the analytically computed results assumed a carrier loop implementation based on an AMP approximation of the transmitted GMSK corresponding to one (the main) pulse stream. For this case, we see virtually perfect agreement between simulation and analytically computed results. For the computer simulation, another option was explored wherein the true GMSK (which requires eight AMP components to fully represent the transmitted waveform) was transmitted. Here, we have a bit of a mismatch between the receiver and the transmitter because the carrier loop is matched to only one of the eight AMP components that compose the GMSK modulation. Thus, at high SNR (where the signal dominates the noise), the simulation reveals a bit of performance degradation. This performance degradation can be diminished by implementing the receiver with a second layer corresponding to the second AMP component and adding the two components to produce the resulting error signal, as was previously suggested. Although this requires additional implementation complexity, in some applications, it may be warranted.

2.9 Simulation Performance

Aside from supporting analysis, simulations are especially useful in providing results in situations where analysis is either unavailable or, because of the complexity of the system model, would be too difficult to perform. In this section, we present some of these simulation results obtained from modeling the various systems on a Signal Processing WorkSystem (SPW) workstation.

Figure 2-47 is a block diagram of the simulation used to model precoded GMSK with concatenated block [(255,233) Reed-Solomon] and convolutional (rate 1/2, constraint length $K = 7$) error correction coding. The uncoded portion of the receiver is based on the suboptimum scheme proposed by Kaleh [46] (and discussed in Sec. 2.8.2.6*b*), which incorporates a Wiener filter following the matched filter prior to the decision device. The idealized (no data imbalance) BEP performance obtained from running this simulation is illustrated in Fig. 2-48, corresponding to values of $BT_b = 0.25$ and $BT_b = 0.50$. Also included for comparison is the performance of BPSK with the same error correction coding. We observe in this figure that whereas coded GMSK with $BT_b = 0.25$ suffers a small E_b/N_0 penalty (relative to coded BPSK) of something less than 0.2 dB, coded GMSK with $BT_b = 0.50$ has virtually identical performance to coded BPSK. This is a rather striking result when one considers the significant improvement in bandwidth efficiency offered by the former relative to the

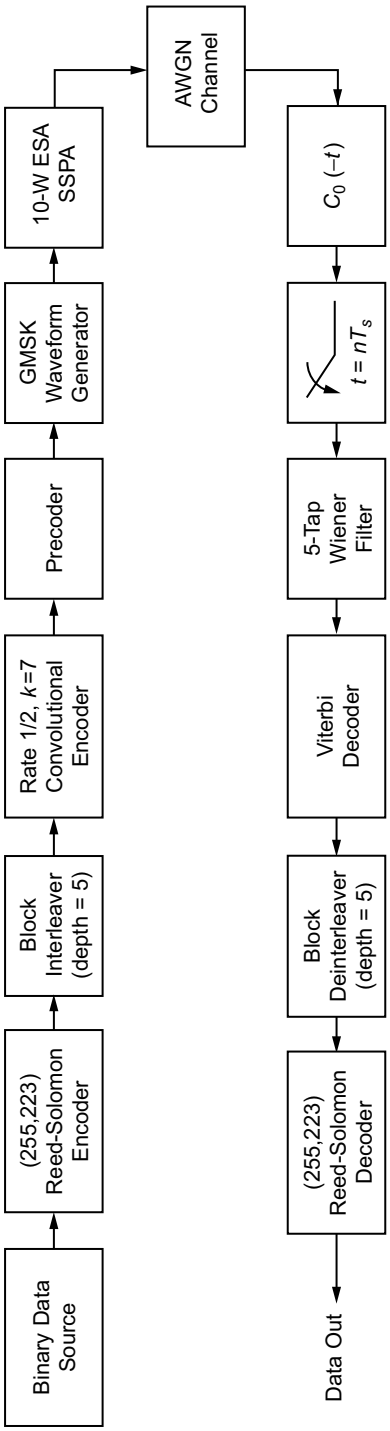

Fig. 2-47. Block diagram of a simulation for precoded GMSK with concatenated Reed-Solomon/Viterbi error-correction coding.

latter. If one now eliminates the error correction coding from the simulation, then equivalent performance results are illustrated in Fig. 2-49. Here again, we observe that GMSK with $BT_b = 0.50$ has virtually identical performance to BPSK. Finally, the performance of uncoded GMSK in the presence of data imbalance is illustrated in Fig. 2-50 for the case of $BT_b = 0.25$. Surprisingly, even with 60 percent data imbalance, the degradation in E_b/N_0 is rather small (on the order of 0.25 dB). If one increases the value of BT_b to 0.5, then even at this rather large data imbalance, the degradation becomes virtually nil. The apparent conclusion to be drawn from what is illustrated in these figures is that while data imbalance has a pronounced effect on the PSD of GMSK, its effect on BEP is quite insignificant.

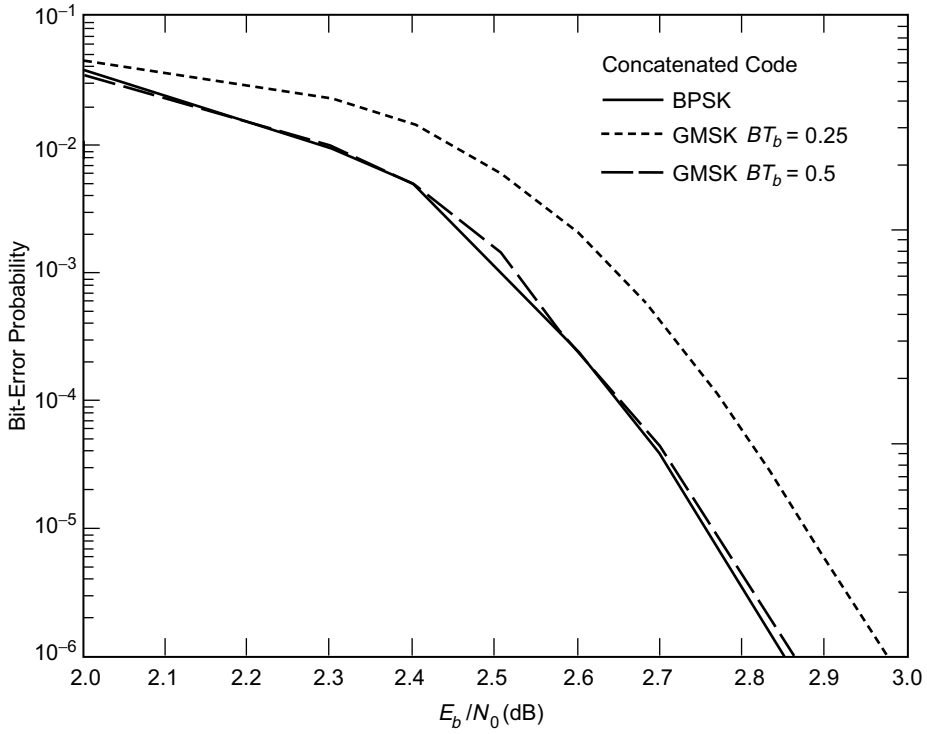

Fig. 2-48. Bit-error probability performance of precoded GMSK with concatenated (Reed-Solomon/convolutional) error-correction coding.

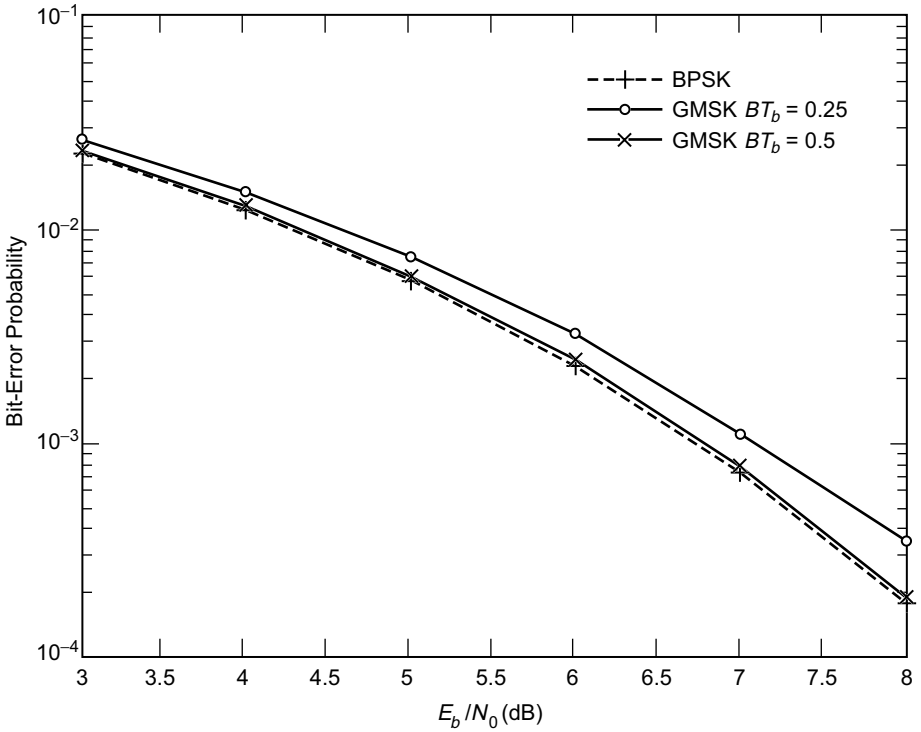

Fig. 2-49. Bit-error probability performance of uncoded GMSK with a suboptimum (Wiener filter-type) receiver.

References

[1] M. K. Simon, S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*, Upper Saddle River, New Jersey: Prentice Hall, 1995.

[2] M. K. Simon and D. Divsalar, "On the optimality of classical coherent receivers of differentially encoded M-PSK," *IEEE Communications Letters*, vol. 1, no. 3, pp. 67–70, May 1997.

[3] T. M. Nguyen, "On the effects of a spacecraft subcarrier unbalanced modulator," *IEEE Journal of Digital and Analog Communication Systems*, vol. 6, pp. 183–192, 1993.

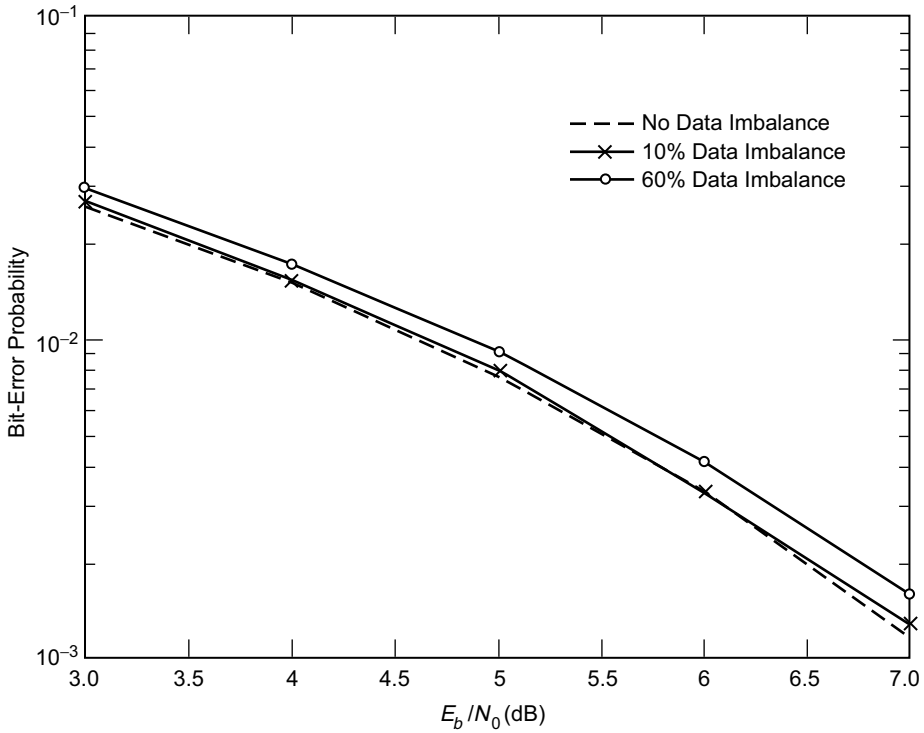

Fig. 2-50. Bit-error probability performance of uncoded GMSK in the presence of data imbalance.

[4] J.-L. Gerner, "A position paper on the effect of phase unbalanced modulator on the performance of PSK modulation schemes for category A missions," Consultative Committee for Space Data Systems, *Proceedings of the CCSDS RF and Modulation Subpanel 1E Meeting*, CCSDS 421.0-G-1, Green Book, pp. 287–301, September 1989.

[5] T. M. Nguyen and Y. Owens, "Cross-talk in QPSK communication systems," Consultative Committee for Space Data Systems, *Proceedings of the CCSDS RF and Modulation Subpanel 1E Meeting*, CCSDS B20.0-Y-1, Yellow Book, pp. 85–93, September 1993.

[6] T. M. Nguyen and A. Anabtawi, "Cross-talk due to phase imbalance between the channels in QPSK communication systems," paper presented at the Consultative Committee for Space Data Systems RF and Modulation Subpanel 1E Meeting, Pasadena, California, June 1994.

[7] M. K. Simon, "The effect of modulator unbalance on QPSK performance," paper presented at the Consultative Committee for Space Data Systems RF and Modulation Subpanel 1E Meeting, Pasadena, California, May 1996.

[8] H. Tsou, "The effect of phase and amplitude imbalance on the performance of BPSK/QPSK communication systems," *Telecommunications and Data Acquisition Progress Report 42-130*, vol. April–June 1997, August 15, 1997, http://tmo.jpl.nasa.gov/progress_report/issues.html Accessed March 2, 2001.

[9] H. Tsou, "The effect of phase and amplitude imbalance on the performance of offset quadrature phase-shift-keyed (OQPSK) communication systems," *Telecommunications and Mission Operations Progress Report 42-135*, vol. July–September 1998, November 15, 1998, http://tmo.jpl.nasa.gov/progress_report/issues.html Accessed March 2, 2001.

[10] H. Tsou, "The combined effect of modulator imbalances and amplifier nonlinearity on the performance of offset quadrature phase-shift-keyed (OQPSK) communication systems," *Telecommunications and Mission Operations Progress Report 42-137*, vol. January–March 1999, May 15, 1999, http://tmo.jpl.nasa.gov/progress_report/issues.html Accessed March 2, 2001.

[11] W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*, Upper Saddle River, New Jersey: PTR Prentice Hall, 1973.

[12] M. K. Simon, "Carrier synchronization of offset quadrature phase-shift keying," *Telecommunications and Mission Operations Progress Report 42-133*, vol. January–March 1998, May 15, 1998, http://tmo.jpl.nasa.gov/progress_report/issues.html Accessed March 2, 2001.

[13] W. B. Davenport, Jr. and W. L. Root, *An Introduction to the Theory of Random Signals and Noise*, New York: McGraw-Hill, 1958.

[14] M. K. Simon, "The effects of residual carrier on Costas loop performance as applied to the Shuttle S-band Uplink," *IEEE Transactions on Communications (Special issue on Space Shuttle Communications and Tracking)*, vol. COM-26, no. 11, pp. 1542–1548, November l978.

[15] J. B. Anderson, T. Aulin, and C.-E. Sundberg, *Digital Phase Modulation*, New York: Plenum Press, 1986.

[16] M. K. Simon, "A generalization of MSK-Type signaling based upon input data symbol pulse shaping," *IEEE Transactions on Communications*, vol. COM-24, no. 8, pp. 845–856, August l976.

[17] M. L. Doelz and E. T. Heald, "Minimum-shift data communication system," U.S. patent no. 2,977,417, March 28, 1961.

[18] P. Galko and S. Pasupathy, "Generalized MSK," *Proceedings of the IEEE International Electrical & Electronics Conference & Exposition*, Toronto, Ontario, Canada, October 5–7, 1981.

[19] I. Korn, "Generalized MSK," *IEEE Transactions on Information Theory*, vol. IT-26, no. 2, pp. 234–238, March 1980.

[20] F. Amoroso, "Pulse and spectrum manipulation in the minimum (frequency) shift keying (MSK) format," *IEEE Transactions on Communications*, vol. COM-24, no. 3, pp. 381–384, March 1976.

[21] M. G. Pelchat, R. C. Davis, and M. B. Luntz, "Coherent demodulation of continuous phase binary FSK signals," *Proceedings of the International Telemetry Conference*, Washington, D.C., 1971.

[22] H. R. Mathwich, J. F. Balcewicz, and M. Hecht, "The effect of tandem band and amplitude limiting on the E_b/N_0 performance of minimum (frequency) shift keying (MSK)," *IEEE Transactions on Communications*, vol. COM-22, no. 10, pp. 1525–1540, October 1974.

[23] S. A. Gronemeyer and A. L. McBride, "MSK and offset QPSK modulation," *IEEE Transactions on Communications*, vol. COM-24, no. 8, pp. 809–820, August 1976.

[24] D. M. Brady, "A constant envelope digital modulation technique for millimeter-wave satellite system," *International Conference on Communications*, Minneapolis, Minnesota, June 17–19, 1974, p. 36C-1.

[25] D. P. Taylor, "A high speed digital modem for experimental work on the communications technology satellite," *Canadian Electrical Engineering Journal*, vol. 2, no. 1, pp. 21–30, 1977.

[26] R. M. Fielding, H. L. Berger, and D. L. Lochhead, "Performance characterization of a high data rate MSK and QPSK channel," *International Conference on Communications*, Chicago, Illinois, pp. 3.2.42–3.2.46, June 12–15, 1977.

[27] B. E. Rimoldi, "A decomposition approach to CPM," *IEEE Transactions on Information Theory*, vol. IT-34, no. 2, pp. 260–270, May 1988.

[28] J. L. Massey, "A generalized formulation of minimum shift keying modulation," *International Conference on Communications*, vol. 2, Seattle, Washington, pp. 26.5.1–26.5.5, June 1980.

[29] T. Masamura, S. Samejima, Y. Morihiro, and H. Fuketa, "Differential detection of MSK with nonredundant error correction," *IEEE Transactions on Communications*, vol. COM-27, no. 6, pp. 912–918, June 1979.

[30] R. DeBuda, "The Fast FSK modulation system," *International Conference on Communications*, Montreal, Canada, pp. 41-25–45-27, June 14–16, 1971.

[31] R. DeBuda, "Coherent demodulation of frequency-shift-keying with low deviation ratio," *IEEE Transactions on Communications*, vol. COM-20, no. 3, pp. 429–435, June 1972.

[32] W. R. Bennett and S. O. Rice, "Spectral density and autocorrelation functions associated with binary frequency shift keying," *Bell System Technical Journal*, vol. 42, no. 5, pp. 2355–2385, September 1963.

[33] R. W. Booth, "An illustration of the MAP estimation method for deriving closed-loop phase tracking topologies: the MSK signal structure," *IEEE Transactions on Communications*, vol. COM-28, no. 8, pp. 1137–1142, August 1980.

[34] S. J. Simmons and P. J. McLane, "Low-complexity carrier tracking decoders for continuous phase modulations," *IEEE Transactions on Communications*, vol. COM-33, no. 12, pp. 1285–1290, December 1985.

[35] J. Huber and W. Liu, "Data-aided synchronization of coherent CPM receivers," *IEEE Transactions on Communications*, vol. 40, no. 1, pp. 178–189, January 1992.

[36] M. Moeneclaey and I. Bruyland, "The joint carrier and symbol synchonizability of continuous phase modulated waveforms," *International Conference on Communications*, vol. 2, Toronto, Canada, pp. 31.5.1–31.5.5, June 1986.

[37] A. N. D'Andrea, U. Mengali, and R. Reggiannini, "A digital approach to clock recovery in generalized minimum shift keying," *IEEE Transactions on Vehicular Technology*, vol. 39, no. 3, pp. 227–234, August 1990.

[38] A. N. D'Andrea, U. Mengali, and M. Morelli, "Multiple phase synchronization in continuous phase modulation," in *Digital Signal Processing 3*, New York: Academic Press, pp. 188–198, 1993.

[39] U. Lambrette and H. Meyr, "Two timing recovery algorithms for MSK," *International Conference on Communications*, New Orleans, Louisiana, vol. 2, pp. 1155–1159, May 1–5, 1994.

[40] A. N. D'Andrea, U. Mengali, and M. Morelli, "Symbol timing estimation with CPM modulation," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1362–1371, October 1996.

[41] K. Murota, K. Kinoshita, and K. Hirade, "Spectrum efficiency of GMSK land mobile radio," *International Conference on Communications*, vol. 2, pp. 23.8.1–23.8.5, June 14–20, 1981.

[42] K. Hirade, K. Murota, and M. Hata, "GMSK transmission performance in land mobile radio," *Global Communications Conference*, pp. B3.4.1-B3.4.6.

[43] K. Daikoku, K. Murota, and K. Momma, "High-speed digital transmission experiments in 920 MHz urban and suburban mobile radio channels," *IEEE Transactions on Vehicular Technology*, vol. VT-31, no. 2, pp. 70–75, May 1982.

[44] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Upper Saddle River, New Jersey: Prentice-Hall, 1996.

[45] A. Linz and A. Hendrickson, "Efficient implementation of an I-Q GMSK modulator," *IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 43, no. 1, pp. 14–23, January 1996.

[46] G. K. Kaleh, "Simple coherent receivers for partial response continuous phase modulation," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 9, pp. 1427–1436, December 1989.

[47] M. R. L. Hodges, "The GSM radio interface," *British Telecom Technological Journal*, vol. 8, no. 2, January 1990.

[48] J. Haspeslagh et al., "A 270 Kb/s 35-mW modulation IC for GSM cellular radio hand held terminals," *IEEE Journal on Solid State Circuits*, vol. 25, no. 12, pp. 1450–1457, December 1990.

[49] R. Hunter and F. Kostedt, "Enhance GMSK performance with two-point modulation," *Microwaves & RF*, vol. 39, no. 4, pp. 59–69, April 2000.

[50] K. Feher, *Wireless Digital Communications*, Upper Saddle River, New Jersey: Prentice Hall, 1995.

[51] F. Wellesplein, "Trends in silicon radio large scale integration," *Microwave Engineering Europe*, pp. 37–45, May 2000.

[52] P. A. Laurent, "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses," *IEEE Transactions on Communications*, vol. COM-34, no. 2, pp. 150–160, February 1986.

[53] U. Mengali and M. Morelli, "Decomposition of M-ary CPM signals into PAM waveforms," *IEEE Transactions on Information Theory*, vol. 41, no. 5, pp. 1265–1275, September 1995.

[54] K. Tsai and G. L. Lui, "Binary GMSK: Characteristics and performance," 99-G1-2, *International Telemetering Conference*, Las Vegas, Nevada, October 25–28, 1999.

[55] G. L. Lui and K. Tsai, "Data-aided symbol time and carrier phase tracking for pre-coded CPM signals," 99-G1-4, *International Telemetering Conference*, Las Vegas, Nevada, October 25–28, 1999.

[56] G. L. Lui and K. Tsai, "Viterbi and serial demodulators for pre-coded binary GMSK," 99-G1-3, *International Telemetering Conference*, Las Vegas, Nevada, October 25–28, 1999.

[57] G. L. Lui, "Threshold detection performance of GMSK signal with BT=0.5," *MILCOM' 98 Conference Proceedings*, vol. 2, pp. 515-519, October 19–21, 1998.

[58] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. IT-13, no. 2, pp. 260–269, April 1967.

[59] J. Proakis, *Digital Communications*, 3rd edition, New York: McGraw-Hill, 1995.

[60] M. K. Simon, P. Arabshahi, L. Lam, and T.-Y. Yan, "Power spectrum of MSK-Type Modulations in the Presence of Data Imbalance," *Telecommunications and Data Acquisition Mission Operations Progress Report 42-134*, vol. April–June 1998, August 15, 1998, http://tmo.jpl.nasa.gov/progress_report/issues.html Accessed March 2, 2001.

[61] L. B. W. Jolley, *Summation of Series*, New York: Dover Publications, 1961.

[62] D. Lee, "Occupied bandwidth of MSK and GMSK in the presence of data imbalance," Consultative Committee for Space Data Systems, *Proceedings of the CCSDS RF and Modulation Subpanel 1E Meeting*, European Space Research and Technology Centre (ESTEC), Noordwiik, The Netherlands, October 18–22, 1999.

[63] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*, New York: Plenum Press, 1997.

[64] S. M. Hinedi, "Carrier Synchronization in Bandlimited Channels," Ph.D. dissertation, University of Southern California, 1987.

[65] M. K. Simon and S. Hinedi, "Suppressed carrier synchronizers for ISI channels," CD-ROM, *Global Telecommunications Conference*, London, England, November 18–22, 1996.

[66] M. K. Simon, "MAP-motivated carrier synchronization of GMSK based on the Laurent AMP representation," CD-ROM, *Global Telecommunications Conference*, Sydney, Australia, November 8–12, 1998.

[67] W. C. Lindsey and M. K. Simon, "Optimum performance of suppressed carrier receivers with Costas loop tracking," *IEEE Transactions on Communications*, vol. COM-25, no. 2, pp. 215–227, February 1977.